ASSET PRICING THEORY AND A COMPARISON OF MACHINE-LEARNING TECHNIQUES

By:

Muhammad Salman Shah

PIDE2018FMPHILETS01

Supervisor: Dr. Mehmood Khalid

Co-supervisor: Dr. Hafsa Hina

DEPARTMENT OF ECONOMICS AND ECONOMETRICS

PAKISTAN INSTITUTE OF DEVELOPMENT ECONOMICS

ISLAMABAD

YEAR: 2021

Pakistan Institute of Development Economics

CERTIFICATE

This is to certify that this thesis entitled: **"Asset Pricing theory and a comparison of machine-learning techniques"** submitted by Mr. Muhammad Salman Shah is accepted in its present form by the Department of Economics & Econometrics, Pakistan Institute of Development Economics (PIDE), Islamabad as satisfying the requirements for partial fulfillment of the degree of **Master of Philosophy in Econometrics.**

External Examiner:

Dr. Mudassar Rashid Khawaja
Assistant Professor
COMSATS University
Islamabad

Supervisor:

Dr. Mahmood Khalid
Senior Research Economist
PIDE, Islamabad

Supervisor:

Dr. Hafsa Hina
Assistant Professor
PIDE, Islamabad

Head, Department of Economics & Econometrics:

Dr. Shujaat Farooq
Head
Department of Economics & Econometrics
PIDE, Islamabad

AUTHOR'S DECLARATION

I Muhammad Salman Shah hereby state that my Mphil thesis titled Asset pricing theory and a comparison of machine-learning techniques is my own work and has not been submitted previously by me for taking any degree from Pakistan Institute of Development Economics or anywhere else in the world. At any time if my statement is found to be incorrect even after my graduation the university has the right to withdraw my Mphil degree.

Date: 5/5/2021.

Signature of student:

Name of student:

Acknowledgements: I am deeply humbled to have this opportunity of writing my thesis on such a diverse topic of asset pricing theory. I am extremely grateful to my supervisors, Dr. Mehmood Khalid and Dr. Hafsa Hina for their invaluable advice, continuous support, and patience during my thesis. Their immense knowledge and experience have motivated me in all time of my academic research. I would also like to thank my fellow researchers for the extensive scholarly discussions and the support staff at Pakistan Institute of Development Economics. Finally, I would like to express my gratitude to God Almighty, my guide, and my parents for the tremendous efforts and encouragement that have offered throughout my life. It would be impossible for me to complete my thesis without their support.

Abstract: The main rationale of asset pricing theory is to identify the underlying pattern of the drivers and establish their relationship with the financial performance of a firm. The proliferation of hundred of drivers often called the curse of dimensionality in the candidate factor pool is a result of continuous research to achieve higher returns than market. Thus, the fundamental task facing the asset pricing theory today is to bring discipline to the zoo of factors. The leverage to utilize machine learning techniques is enhanced due to their innate ability of handling and extracting valid signals from such complex data structures. The most notable techniques are the neural networks, tree based models, and penalized regressions. The neural networks on the training sample of US market performed the best with the MSE of 3% and hit ratio of 55%. The most prevalent factors include the market capitalization confirming the existence of size anomaly, momentum indicators, and the capital expenditure to sales cash flow ratio among others. The MSE and the hit ratio for the data of Pakistan, where the best performing candidate model is random forest is 0.3% and 84% respectively. Most significant contributors for the data of Pakistan includes momentum, price volatility, and dividend yield etc. The results are encouraging but still warrants further research especially to the formulation of ensembles that may beat the naive equal weighted ensembles.

Abbreviations:

Table 1: Abbreviations used in the thesis

United States	US
Pakistan	Pk
Market Capitalization	Market Cap
Price to book ratio	Pb ratio
Price to sales ratio	Ps ratio
Mean Absolute error	MAE
Mean Squared error	MSE

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Background

1.1 Introduction

The capital market is the barometer of the economy by which we can assess and guarantee the sustainable growth in an economy. The issuance of shares as a mean to raise new capital is significant source of financing available to companies. Equity markets are relatively more flexible in offering greater variety of financing options as compared to debt markets. The equity instruments most commonly traded in the form of common stock in a stock market. In addition, the stock market also provides an opportunity for an ordinary individuals to invest their savings in attractive securities with higher returns.This goal of capital formation is important in promoting the level of savings and enabling the country to efficiently allocate the resources to achieve economic growth.

Efficient markets may be desirable for the society, as prices determine the allocation of economic resources. Most of the early work that relates to efficiency theory stems from closely related concept of random walk hypothesis (RWH). The RWH asserted that the changes in the markets are completely random and hence unpredictable, later consolidated, by (Fama, 1965) in form of efficient market hypothesis. The (EMH) has stirred both the interest and controversial views among the academics and practitioners alike, which resulted in staggering evidence both for and against the efficiency of market prices. The support for EMH came from the growing body of empirical research that demonstrated the difficulty of beating the average return in the stock market index

((Fama & French, 2015). On the contrary, studies that describe deviations from the EMH are termed as anomalies, are the biggest threat to the hypothesis of efficient markets. Asset pricing anomalies, whether they are broad macro-economic factors or company specific characteristics became the foundation of asset pricing theory.

The practitioners aligned their investment strategies with groups of stocks, such as value stock, growth stock, or other stock metrics such as income. The focus is to earn positive alpha returns that usually aims to exploit the short-term mis-pricing. Academic research on using the risk factors in asset pricing dates back to the seminal study of (Graham & Dodd, 1934), where he wrote about value premium. The researchers dig deeper to explain the underlying phenomenon of the equity markets and as a result two equilibrium models originated to map the relationship between asset returns and factors: the capital asset pricing model (CAPM) and the arbitrage pricing theory (APT). The CAPM provided an important foundational principle of modern finance, stating that the security return is a function of market risk. However, APT demonstrated that the returns of securities can be modeled as a function of various risk factors. The APT provided an empirical based framework, launching the tradition of using multiple factors as a research tool, to understand the risk and return characteristics of different securities.

The two most commonly used set of factors to analyze and predict stock returns are fundamental and technical analysis. The former approach studies the general economic and company-specific factors, and is best suited for a longer-term prediction spectrum. On the other hand, the technicians suggest that prices already reflect all the fundamental information and historical behavior of financial asset tend to repeat itself (Murphy, 1999). But as noted earlier, the financial series are essentially a complex phenomenon and the real-time aspect of financial prediction further adds an additional layer of complexity as the temporal dimension, where the influence of time is different for different attributes and vary across the cross-section of asset classes. This makes the predictability of stock returns in both the time series and cross-section context a challenging task as it needs to account for a myriad of confounding factors including both technical and fundamental factors.

Several academic disciplines including finance, mathematics, and engineering have tried to model and predict the behavior of equity markets (Yoo, Kim & Jan, 2005). Traditionally, statistical models and econometric techniques were employed to forecast equity returns, which generally assume that the equity returns are generated through a specified underlying probability distribution. (Kumar & Murugan, 2013). However, the time series of equity returns is essentially very complex, highly noisy, dynamic, non-linear, non-parametric, and chaotic in nature (Si & Yin, 2013).

Recently the advent of computer science methods provided leverage to apply advance data analytics algorithms to extract valid signals from such a highly complex financial data. In other words, machine learning seeks to extract knowledge from large amounts of data without any restrictions. The goal of machine learning algorithms is to automate decision-making processes by generalizing from known examples to determine an underlying structure in the data. The emphasis is on the ability of the algorithm to uncover structure or predictions from data without any human help.

The success of machine learning in asset pricing has it source at the convergence of three favorable developments: data availability, computational capacity, and economic groundings. In order to model and make predictions about the future expected returns, machine learning techniques have been applied with relative success (Lee, 2009). Machine-learning techniques are better able than statistical approaches to handle problems with many variables (high dimensionality) or with a high degree of non- linearity. ML algorithms are particularly good at detecting change, even in highly non- linear systems, because they can detect the preconditions of a model's break or anticipate the probability of a regime switch. (Atsalakis & Valavanis, 2009). As a result, financial models became more capable of dealing with real-time prediction problem of stock prices.

Asset-pricing models that use machine-learning techniques is an emerging topic, offering both theoretical and practical perspectives to equity trading and is at the cross-section of finance, econometrics and computer science. The theoretical underpinning is provided by the field of finance and the econometrics and computer science provides the analytical tools to operationalize financial theories and models. As a results, we yield intelligent decision support and prediction models that are of practical use in understanding the drivers of equity returns. Reliable and action-able financial prediction is still a challenge given the demand for real-time forecasting, in the presence of a myriad of factors that are probably interlinked with and influenced by geopolitical and socio-economic phenomenon.

The chapter 2 includes the background, objectives, and the problem statement. Chapter 3 provides with a comprehensive literature review of both the evolution of factors in asset pricing theory and the application of machine-learning tools. Chapter 4 introduces with the data acquisition process and the pre-processing. In this chapter, we also lay down the methodology, which we will be utilizing to gather and interpret results. Chapter 5 illustrates and discusses the results. Chapter 6 is related to building and testing the different ensemble models. Finally, chapter 7 discusses the outcome and the future directions.

1.2 Background, Objectives, and Statement of problem

1.2.1 Background

(Bachelier, 1900) with his seminal contribution in 'The Theory of Speculation' deeply influenced the whole development of mathematics of equity pricing theory. Bachelier's work wasn't applied to Economics until the beginning of modern financial theory in 1965, when (Samuelson, 2016) demonstrated that in a well-informed and competitive speculative market, asset prices will be random. Samuelson's proof together with his further research in collaboration with (Fama, 1965) has been crucial to the development of the efficient market hypothesis (EMH). The EMH postulated that the prices reflect all available information i.e. technical indicators, fundamental factors and insider information in the stock prices and thus, only new information explains the movement of stock prices. Since the new information comes in a random way, thus the stock prices also fluctuate randomly making the predictability of stock returns infeasible.

For every research producing empirical evidence in favor of market efficiency, a contradictory paper is available to establish market inefficiency, which is often termed as anomaly (Majumder, 2013). Anyone who has observed price behavior following an earnings surprise, in which price movement has been excessive and would be expected to self-correct within a few trading days. This tendency is far from random and is caused by a variety of distortions in supply and demand. Thus, the belief in EMH is questioned because it claims that market participants will perform at the market average over time. Many characteristics, for example price momentum, price earnings ratio, dividend yield etc. clearly proves the postulates of EMH wrong.

The two main approaches commonly used to analyze and predict financial market behaviors are the technical analysis and the fundamental analysis. The technical analysts like the practitioners tend to exploit the short-term inefficiencies and believe that the movement in the stock price and volume reflects all the relevant information. Moreover, technicians assume that these past trends in prices and volume tend to repeat themselves in the future, thus by identifying the previous behavior patterns, trading rules can be formulated to achieve higher returns (Esfahanipour & Mousavi, 2011).

It is important to note that predictability and earning higher abnormal returns using technical analysis are only possible if weak form EMH is rejected i.e. historical prices do not reflect all past information in future. Regardless of the theories supporting EMH in general and weak form in specific, investors, financial experts and brokerage firms have used technical analysis with considerable success (Rodríguez-González, García-Crespo, Colomo-Palacios, Iglesias & Gómez-Berbís, 2011). Several scientific papers in the literature have proposed many technical indicators that explain the variability in stock market returns e.g. (Aghabozorgi & Teh, 2014). The most persistent technical anomalies that poses serious questions to market efficiency are momentum or trend following effects (Luo, Subrahmanyam & Titman, 2019) and emergence of bubbles and market crashes.

Fundamentalists closely knitted with the academics, on the other hand, believe that the stock prices of the company is reflected by several political and economic factors that are internal and external to the company. This belief is in contrast with the semi-strong efficient market hypothesis, which postulates that all available information is already reflected in equity prices.

Studies that examined the results of corporate insiders and stock exchange specialists (Schwert, 2003) suggests that both corporate insiders having insider information and stock exchange specialists having better analytical skills and tools earn abnormal returns by beating the world most sophisticated and informationally strong-form efficient stock markets consistently. Few examples of fundamental anomalies are returns based on price earning ratios, size, neglected stocks, book value to market value ratios, outperformance of value stocks in favor of growth stocks (Richardson, Tuna & Wysocki, 2010) among others. This evidence contradicts the semi-strong hypothesis since all these fundamental factors are public knowledge, and this contradictory evidence suggests that some combination of technical and fundamental analysis can predict the market.

Furthermore, an additional layer of complexity is added when (Malkiel & Fama, 1970) demonstrated that the hypothesis of efficient markets could not be rejected without an accompanying rejection of the model i.e. model misspecification. This difference between the theory and the reality, known as a joint hypothesis problem and we can only test a joint hypothesis by stating that the market is efficient in equating asset prices with their intrinsic values, where the intrinsic values are determined by a perfect asset pricing model. Hence, whenever an anomaly is found, we have no way of knowing that which part of this joint hypothesis did not work. Most studies that deny the existence of anomalies shift the blame towards incomplete models of stock prediction. In the words of (Fama, 1998) inferences about market efficiency can be sensitive to the assumed model for expected returns.

1.2.2 Research Questions

The main rationale behind the asset pricing theory is to identify the underlying phenomenon in the volatility of asset returns. Thus, anomalies or factors that explain asset prices are considered to be a corner-stone of factor models in asset pricing literature. In other words, the core objective of a factor model is to identify the underlying patterns and establish a relationship of the identified drivers with the financial performance of a firm. This subject is incredibly large and there are many papers dedicated to it and it is still growing with a fast pace. So naturally, the first essential question is to identify the relevant drivers of asset returns. In answering the first question, we will establish that drivers are either the individual firm characteristics or broad macro-based factors, which is done in the literature review presented in Chapter 2 of this study. The sequential literature review will help us identify the most common drivers that explain the cross-section of asset returns. There are hundreds of drivers already identified in the literature as explained in the background study and the candidate pool is still evolving both in the sense of quantity and dynamism amongst the variables. Thus, the fundamental question is to identify the most relevant drivers of asset returns in such a high-dimensional, non-linear, complex, chaotic, and dynamic space.

The machine-learning techniques offer more adaptive and flexible mechanisms in optimal feature selection, as they are well suited to handle high-dimensional, non-linear, and complex spaces ((Lee, 2009). Specifically, penalized regressions, tree based models, and neural networks are amongst the famous feature-selection techniques that are well suited in high-dimensional space. In this light, we will present our methodology in section 3 and sequentially test the question of in-sample and out-of-sample explanatory predictive power of the techniques we are employing, which is discussed in section 4 and 5. We will also evaluate the feasibility of different ensembles based on the combination of techniques that are employed. Since, machine-learning techniques are hard to interpret, we finally address the question of explaining how the respective decision is being made in each network.

1.2.3 Objectives of the study

- Identify and investigate the formal representation of key underlying factors or characteristics that are determinants of stock returns in the presence of hundreds of potential factors using machine-learning tools.
- Evaluation and comparison of machine-learning techniques and ensembles in identifying and explaining the underlying factors of stock returns. The tools that we used are shrinkage techniques including Ridge regression, Lasso, and Elastic Net, tree-based techniques including simple-trees and random forests, and neural networks.
- Opening up the traditional black-boxes by ascertaining the feature importance of each algorithm.
- The ensembles are formed utilizing two completely different markets i.e. the US and Pakistan in order to evaluate any potential diversification benefits. We will further provide a comparison of the predictive out-of-sample performance of the techniques used.

1.2.4 Significance of the study

The reason on which the factor investing is based is to identify the relationship between the financial performance of a firm and underlying drivers. Explaining the underlying phenomenon of asset returns is the fundamental objective of any factor model and the search for factors has produced a large amount of potential candidates. The fundamental task facing the asset pricing theory today is to bring more discipline to the proliferation of factors (Cochrane, 2009).

Most of the studies that explain the variability of stock returns are concentrated to employ only past price and volume information also called technical indicators (Cavalcante, Brasileiro, Souza, Nóbrega & Oliveira, 2016). Other studies such as fama-five factor model (Fama & French, 2015) focuses on the factors of portfolios that are usually a long position in the best performing stock and short position on the lowest quantile stocks rather than just the pure characteristics of firms. There are very few recent studies discussed in the literature review in the US market, while we do not find any related study in the case of Pakistan. Thus, the main contribution of our study is the advancement of asset pricing theory through a lens of factor investing in the field of Finance.

This thesis aims to contribute the subject of asset pricing by addressing the problem of extracting a valid signal from a high-dimensional stock market space. The answers to the research questions presented above are exploratory and expected to alleviate the uncertainty among the participants of the equity markets by providing them with a comparative performance of recent machine-learning techniques in selection of a robust model. The process of model building is unique to this study, as this study mainly utilizes the individual level firm characteristics rather than just relying on the raw price and volume data to build up the candidate feature library. The feature space for the data of US is 93 variables for 1200 stocks and 40 variables for all 19 major banks of Pakistan. To best of our knowledge, there is no other study specially in the case of Pakistan that utilizes such a diverse feature space to extract relevant signals that explain the cross-section of different firms.The most closely related study is (Jan, 2019), but the scope of this study is significantly different in many ways. Primarily (Jan, 2019) dealt with the forecasting ability of CAPM and fama-french factor models using 6 indicators in context of artificial neural networks. On the contrary, we take an approach to value a firm based on individual 40 characteristics. Moreover, we have comprehensively compared the performance and built ensembles on penalized regressions, tree based models, and neural networks.

Additionally, the factors are extracted by employing the most recent and cutting-edge machine-learning techniques, so we are also contributing on the application of machine-learning techniques in the context of asset pricing. More specifically, our contribution to the machine-learning literature is geared on the comparison of different techniques and their respective explanatory power in a high-dimensional setting for robust model building. As machine-learning techniques are generally considered as black-boxes, the revelation of the feature importance under each respective network is an important contribution of this study. Lastly, we will build ensemble of different models on two completely different datasets to maximize the diversification benefits from the models.

Chapter 2

Literature Review

Financial markets are the drivers of modern market economy and they reflect the expected growth prospects and risk associated with the firms. It also implies that investors can get significant amount of information from these risk factors to diversify their losses and to perform investment decision-making. In last few decades, various risk factors were identified and reported. The problem is that those risk factors do not persist and continuously evolve over time (Shanaev & Ghimire, 2020). This is the reason of the proliferation of factors. (Cochrane, 2011) labeled the current state of research a ”factor zoo” highlighting that many papers over the past three decades have reported various factors providing excess risk adjusted returns, but they remain not relevant over time. Thus, any reasonable model of prediction must be dynamic enough to cope up with ever changing nature of high dimensional variables

The section 3.1 provides a literature review of the evolution of modeling techniques and the meta studies covering anomalies. This section further signifies the importance of characteristics in comparison to the long-short portfolio based factors. In section 3.2 onward, we present a systematic review of the studies related to the application of machine-learning tools with more focus on predicting the stock returns.

2.1 Evolution of asset pricing models and anomalies

Stock valuation models are a significant tool as they enable a researcher to value the shares in presence of risk factors. Foremost is the modern portfolio theory by (Markowitz, 1952) is referred as one of the most important studies in financial economics literature. The theory proposed that investors should focus on selecting portfolios based on their overall risk-reward characteristics. Modern portfolio theory is the basis for developing the theories of price formation for financial assets, the famous Capital Asset Pricing Model (CAPM) (Sharpe, 1964) and arbitrage pricing theory (APT) Ross(1976). CAPM is perhaps the most significant study in the field of modern financial economics, and it has long been a guide for academics and practitioners interested to model the relationship between average returns and market risk.

CAPM is a market equilibrium model in which market βs is the only relevant measure of an asset's risk. Hence, the cross-section of asset returns only depend on the cross-section of market β. This approach was criticized by (Jensen, Black & Scholes, 1972), which argued that given the condition of high volatility of market returns, β underestimates the overall risk. To address the cross-sectional problem, (Fama & MacBeth, 1973) provided a solution as Fama-Macbeth cross-sectional regression.

The intense econometric investigation of CAPM also lead to the development of several different versions on much more realistic assumptions. The foremost among them was the (Fama & French, 1992) three-factor model. The three-factor model extended the basic CAPM to include size and book-to-market as explanatory variables in explaining the cross-section of stock returns.

After rigorous testing, the three-factor model could not explain the momentum effect presented by (Jegadeesh & Titman, 1993), which lead to the development of four-factor model by (Carhart, 1997). However, the latest studies have proved that four-factor models are inadequate to explain the stock pricing and returns. Recently, (Fama & French, 2015) presented a five-factor model, which adds profitability and investment in addition to the factors from three-factor model. The launch of five-factor model had proved to be an enormous improvement compared to the previous models. However, it has left opportunities for better models to be further developed in future (Rowshandel, Anvary Rostamy, Noravesh & Darabi, 2017) and hence, the search for a better asset pricing model is still on.

The meta-studies that document the historical anomalies include (Green, Hand & Zhang, 2013), (Harvey, Liu & Zhu, 2016), and (McLean & Pontiff, 2016) among others. However, a few most frequently cited factors and their time-lines are summarized below:

- Size anomaly (SML): That explains the out performance of small cap firms in comparison to the large cap firms: (Astakhov, Havranek & Novak, 2019), (C. Asness, Frazzini, Israel, Moskowitz & Pedersen, 2018), (Van Dijk, 2011), (Fama & French, 1993), (Fama & French, 1992), (Banz, 1981).
- Value anomaly (HML): Out performance of value stocks as apposed to the growth stocks:(C. S. Asness, Moskowitz & Pedersen, 2013), (Fama & French, 1993), (Fama & French, 1992).
- Momentum (WML): An anomaly where stocks having highest average returns continue to out perform those with weak returns: (C. S. Asness et al., 2013), (Carhart, 1997), (Jegadeesh & Titman, 1993).
- Profitability (RMW): Firms with strong profitability surpass the returns of firms with weaker profits. (Bouchaud, Krueger, Landier & Thesmar, 2019), (Fama & French, 2015),
- Investment (CMA): Firms with aggressive investment philosophy exceed the returns in comparison to the conservative investments. (Hou et al., n.d.), (Fama & French, 2015).
- Betting against beta (BAB): The strategy predicts that the assets with higher beta are overvalued and assets with lower beta are under-valued and both are mean reverting. (C. Asness, Frazzini, Gormsen & Pedersen, 2020), (Baker, Hoeyer & Wurgler, 2019), (Boloorforoosh, Christoffersen, Fournier & Gouriéroux, 2020), (Frazzini & Pedersen, 2014), (Baker, Bradley & Wurgler, 2011), (Ang, Hodrick, Xing & Zhang, 2005).

It is often convenient to look into the returns of portfolio instead of individual stocks, as portfolios are more stable and have desirable statistical properties. It is easier to detect anomalies from a long-short combination of highest quantile portfolio minus the lowest quantile extreme. It is a current topic in the academic debate of the asset pricing literature, that whether the firm returns are explained by the exposure to the long-short factor portfolios or simply by the pure characteristics of the firms. (Daniel & Titman, 1997) along with two subsequent papers (Daniel, Titman & Wei, 2001) and (Daniel, Titman et al., 2012)) provided evidence in favor of individual firm level characteristics. Small market cap firms with high book-to-market ratios consistently out-performed the average returns, even if other factors are not very positive for them. Therefore, the significance of intrinsic characteristics cannot be ignored against the long-short factor exposure. Some earlier contributions that are made to explain and predict returns with firm attributes include (Hjalmarsson & Manchev, 2012), (Ammann, Coqueret & Schade, 2016), (DeMiguel, Martin-Utrera, Nogales & Uppal, 2020), and (McGee & Olmo, 2020), but these studies are originally not with the intent or focus of machine-learning perspective. The role of characteristics in explaining the return variation with the machine-learning perspective are:

- (Chordia, Goyal & Shanken, 2017) reiteriates the significance of characteristics in comparison to the factor loadings and in their sample characteristics based models have more explanatory power in the variation of expected returns in comparison to the factor-based models.

- (Kozak, Nagel & Santosh, 2018) took a different approach by incorporating the sentiments based risk premium approach for model building.

- (Han, He, Rapach & Zhou, 2019) employed penalized regressions to predict the monthly returns of US stock and they extracted more than 20 characteristics from the pool of 90 variables.

- (Kelly, Pruitt & Su, 2019) and (S. Kim, Korajczyk & Neuhierl, 2019) took an approach by combining both the factors and characteristics by taking factors as latent and beta loadings on characteristics.

- (Koijen & Yogo, 2019) have proposed a demand model in which characteristics are utilized to form the porfolios. They demonstrated that aggregate demand is directly linked to characteristics and not to the factors. In subsequent studies, (Koijen, Richmond & Yogo, 2019) show that only a few sets of characteristics has large explanatory power in the prediction of future returns. Finally, (Martin & Nagel, 2019) proposed that the characteristics play important role in the predictability of dividend growth, which in turn explains the variability of stock returns.

2.2 Machine Learning

Due to the development of computing power and availability of data, machine learning techniques offered a huge improvement in comparison to traditional statistical methods. Machine learning techniques have been applied with relative success in modeling and predicting the asset returns (Lee, 2009). Many machine learning techniques are able to extract nonlinear relationship between the factors without any prior information (Atsalakis & Valavanis, 2009). Most of the applications in the realm of machine learning in asset pricing uses only technical indicators, which might be motivated by the fact that technical indicators are reported daily and easily available. However, we present a case of using both fundamentals and technical indicators in machine learning context in this thesis.

2.2.1 Penalized Regression

We introduce the widespread concept of regularization for linear models. There are several possible applications for these models. The first one is to resort to penalization to improve the robustness of factor-based predictive regressions. For instance, (Han et al., 2019) and (Rapach & Zhou, 2020) used penalized regression to improve stock returns predictions that emanates from individual firm characteristics. Similar ideas are also developed for macroeconomic predictions (Uematsu & Tanaka, 2019). The second application stems from (Stevens, 1998), where he links the weight of optimal mean-variance portfolios to particular cross-sectional regressions. The idea is to improve the quality of mean-variance driven portfolio weights. In any case, the idea presented in seminal paper of (Tibshirani, 1996) is same: standard unconstrained optimization programs may lead to noisy estimates, thus adding a structuring constraint helps remove some noise. For instance, (Kremer, Lee, Bogdan & Paterlini, 2020) use this concept to build more robust mean-variance (Markowitz, 1952) portfolios and (Freyberger, Neuhierl & Weber, 2020) use it to single out the characteristics that help explain the cross-section of equity returns. The focus of this paper is, however, first objective of improving the robustness of factor-based predictive regressions.

2.2.2 Tree-based Models

After the monograph of (Breiman, Friedman, Stone & Olshen, 1984) popularized the powerful yet simple clustering algorithms in form of classification and regression trees. Recently, the surge in Machine Learning applications in Finance has led to multiple publications that use trees in portfolio allocation problems. A list includes (Ballings, Van den Poel, Hespeels & Gryp, 2015), (Patel, Shah, Thakkar & Kotecha, 2015), (Moritz & Zimmermann, 2016), (Krauss, Do & Huck, 2017), (Gu, Kelly & Xiu, 2020), (Guida & Coqueret, 2018), (Coqueret & Guida, 2020b) and (Simonian, Wu, Itano & Narayanam, 2019). One notable contribution is (Bryzgalova, Pelger & Zhu, 2019) in which authors create factors from trees by sorting portfolios via simple trees, why they call Asset Pricing Trees.

2.2.3 Neural Networks

Neural networks (NNs) are a very rich and complex subject. We refer to the definition of NNs given by (Francois, 2017) "chains of differentiable, parameterised geometric functions, trained with gradient descent". (Bansal & Viswanathan, 1993) and (Eakins, Stansell & Buck, 1998) are few of the early adapters of neural networks in financial economics. (Bansal & Viswanathan, 1993) utilized a pricing kernel by estimating nonlinear functional form of the model. While scope of (Eakins et al., 1998) is to identify and quantify the relationship between institutional investments and the attributes of the firm. (Burrell & Folarin, 1997) provided with an early review of financial applications of NNs during the 1990s, while more recently, (Sezer, Gudelek & Ozbayoglu, 2020), (Jiang, 2020) and (Lim & Zohren, 2020) surveyed the deep learning models that attempt to forecast financial time series. Since the predictive out-performance of NNs in financial markets is a popular subject and we further cite these recent additions by (Krauss et al., 2017), (Fischer & Krauss, 2018), (Aldridge & Avellaneda, 2019) and (Soleymani & Paquet, 2020) for interested reader. Below we provide a brief description of some recent studies on neural networks in the realm of financial economics:

- (Feng, Polson & Xu, 2019) studies the factors explaining the stock returns in cross-sectional data emplying neural networks.

- (Gu et al., 2020) studies the relationship between macro-economic variables and firm characteristics to forecast future returns.
- (Chen, Pelger & Zhu, 2019) employed a deep neural network to estimate the pricing kernel.

2.2.4 Model Comparison

Earlier studies that made a comparison between the machine-learning techniques include (Kim, 2003), (Huang, Nakamori & Wang, 2005),(Matias & Reboredo, 2012),(Dunis, Likothanassis, Karathanasopoulos, Sermpinis & Theofilatos, 2013), and (Gu, Kelly & Xiu, 2021). These studies are based on only on daily price and volume information. However, (Guida & Coqueret, 2018) and (Tobek & Hronec, 2020) are the most recent studies that incorporate a large-cross section of characteristics and utilizes machine-learning techniques. (Guida & Coqueret, 2018) is not with the foucs on the comparitive performance of the techniques, rather they utilized aggregated anomalies to create one mis-pricing signal. (Tobek & Hronec, 2020) is the most closely related study, which has drawn a comparison on random forest, boosted trees, and neural networks. (Tobek & Hronec, 2020) has taken it as a classification task. Further, they have not taken penalized regression and regression trees into account.

2.3 Conclusion

The existence of hundreds of potential factors explaining the expected equity returns as noted by (Cochrane, 2011) and more recently by (Harvey et al., 2016), (McLean & Pontiff, 2016), and (Hou, Xue & Zhang, 2020) aspires a researcher naturally of the search for factors that contribute to explain the asset returns. The key findings of the literature review is the evolution of risk factors and factor models is a continuous process. More than 300 risk factors are identified by the researchers, but only few of these essential risk factors are significantly responsible in explaining the variation of stock market returns. Moreover, due to continuous evolution and changing nature of risk factors, the essential risk factors may lose their efficiency in the future. Thus, it seems quite difficult to have a stable efficient factor model that can capture the stock market risk and return relationship globally in long run.

The infusion of evolving factors in the asset pricing space is enormous, but they also provide a unique opportunity to test and compare the performance of new emerging machine learning techniques. Thus, the fundamental task facing the asset pricing field today in the words of (Feng, Giglio & Xiu, 2020) is to bring more discipline to the proliferation of factors. To tame the zoo of factors, machine-learning techniques can leverage the non-linear pattern recognition abilities even in high-dimensional space. There is a lot of space available for this current research especially if we take the problem as regression exercise not classification as demonstrated in the literature review. This will allow us to draw comparison amongst diverse set of techniques, which is not available in the literature.

Chapter 3

Methodology and Data

3.1 Methodology

3.1.1 Introduction

As already explained above, there are many known determinants of stock market returns and the relationship keeps on evolving over passage of time. In such a high-dimensional data with a large number of independent variable, the standard regression models does not suffice the purpose of our prediction. The strong presence of multicollinearity reduces the precision of estimates, thus weakening the statistical power the regression models. Standard p-values are not trustworthy to identify the independent variable that are statistically significant. In order to extract valid signals, resorting to some sort of dimension reduction technique must be used. Traditional techniques include step-wise regressions and information criterion, but they lack in their flexibility. The advent of machine learning techniques has now made it possible to apply some sort of penalization without relying on the restrictive assumptions. The ability of the machine-learning techniques to identify patterns especially the non-linear interaction of variables is unparalleled. Thus, in this thesis, we will resort to supervised machine-learning to extract the relevant characteristics and forecast returns in the cross-section of different firms. The baseline equation in supervised learning is,

$$\mathbf{y} = f(\mathbf{X}) + \epsilon, \tag{3.1.1}$$

In financial terms, the equation can be expressed as

$$\mathbf{r}_{t+1,n} = f(\mathbf{x}_{t,n}) + \epsilon_{t+1,n}, \tag{3.1.2}$$

where $f(\mathbf{x}_{t,n})$ can be viewed as the expected returns for time $t + 1$ computed at time t, which is $\mathbb{E}_t[r_{t+1,n}]$. Building a fair model for predictions or inference requires to delve in all parts of $eq.2$. Sequentially the first step is gathering and processing of the data, which will be discussed later. On the right side of the equation, there is a consensus in the literature as discussed above, the features should include classical predictors such as market capitalization, accounting ratios, risk measures, and momentum proxies. For the dependent variable, most researchers work with monthly returns, but there exists the possibility that other maturities may also perform better in out-of-sample performance. While the significance of the choice of the model f is a crucial part, but the selection and engineering of the inputs are at least as important. Finally, $\epsilon_{t+1,n}$ the errors are most commonly dealt with vanilla quadratic programming. Theoretically, arbitrage pricing theory (APT) of Ross (1976) explains all of the linear factor models to be a special cases of this theory. The return on assets is assumed to be a linear combination factors under APT. Mathematically,

$$r_{t,n} = \alpha_n + \sum_{k=1}^{K} \beta_{n,k} f_{t,k} + \epsilon_{t,n}, \tag{3.1.3}$$

where $\mathbb{E}[\epsilon_{t,n}] = 0$, $\text{cov}(\epsilon_{t,n}, \epsilon_{t,m}) = 0$ and $\text{cov}(\mathbf{f}_n, \boldsymbol{\epsilon}_n) = 0$. If possibility of such factors is proven, then it is in contradiction with the capital asset pricing model (CAPM)(Jensen et al., 1972), according to which market risk is the only driver of asset returns. Restricing the model to the linear relationships of only market risk and prices is not plausible beyond simple interoperability of the CAPM model. We will extend the eq. (2) to serve the dual purpose of capturing non-linearities and prediction as,

$$r_{t+1,n} = g(\mathbf{x}_{t,n}) + \epsilon_{t+1}. \tag{3.1.4}$$

The most obvious difference between (4) and (2) is the introduction of the nonlinear function g. The second difference between the equations is the shift of time index. The interest in the prediction of stock prices in the cross-section of firms will provide a forecast of future average returns. Once the nonlinear model $\hat{g}$ is established, which is the time-t measurable value $g(\mathbf{x}_{t,n})$, it will map the relationship between the predictors and the interest variable. There are several other specifications, in which machine-learning related equations could be used to estimate asset pricing models, beyond the explicit form of Equation (4). One mainstream way to price assets is by stochastic discount factor (SDF) M_t, which satisfies $\mathbb{E}_t[M_{t+1}(r_{t+1,n} - r_{t+1,f})] = 0$ for an asset n (Cochrane, 2009). Another alternative method is to model asset returns as linear combination of factors, just as in (2) and (3). Allowing the loadings $\boldsymbol{\beta}_{t,n}$ to be time dependent, the compact form becomes,

$$r_{t,n} = \alpha_n + \boldsymbol{\beta}'_{t,n}\mathbf{f}_t + \epsilon_{t,n},$$

An important theme is to introduce the independent variables as firm characteristics in the equation above. As also in (Fama & French, 1993) study, characteristics are traditionally present in the definition of factors. The portfolios are constructed on basis of some characteristics like market size, accounting ratios, past performance etc to identify the factors that explain the variability of stock returns. The factors may be constructed heuristically in the portfolios from simple rules like thresholding or sorting. For example, firms within the lowest quantile from book-to-market perspective may be classified as growth stocks and those in the upper quantile as value firms. So a long-short portfolio of the lowest quantile and the highest quantile can be defined as a value factor. It is important to note that (Fama & French, 1993) used a more advanced approach to build the value factor that also takes market capitalization into account. More recently, the process of the construction of factors is automated by the advances enabled by machine learning. One such application is (Feng et al., 2019), where the authors developed a way to optimize the automatic construction of factors for better fit in the cross-section of different assets. Theoretically such models are more generalized and thus the resultant factors help explain a greater proportion of in-sample variation in stock returns.

A third approach was put forward by (Kelly et al., 2019), where the beta loadings are determined by the characteristics and factors are considered unobservable. Naturally, this approach suffers with degrees of freedom problem because in $r_{t,n} = \alpha_n + (\boldsymbol{\beta}_{t,n}(\mathbf{x}_{t-1,n}))'\mathbf{f}_t + \epsilon_{t,n}$, only characteristics $\mathbf{x}_{t-1,n}$ are well defined, while both the factors $\mathbf{f}_t$ and the functional form $\boldsymbol{\beta}_{t,n}(\cdot)$has to be estimated. But the major advantage of Kelly, Pruitt, and Su (2019) is their linear functional form of the model, which is naturally more tractable. Lastly, (Gu et al., 2020) introduced another approach, which combines two neural network architectures. The first neural network processes the characteristics $\mathbf{x}_{t-1}$ as inputs and generate factor loadings $\boldsymbol{\beta}_{t-1}(\mathbf{x}_{t-1})$. A second neural network is then used to transform the returns $\mathbf{r}_t$ into factor values $\mathbf{f}_t(\mathbf{r}_t)$. The aggregate model can then be written as:

$$\mathbf{r}_t = \boldsymbol{\beta}_{t-1}(\mathbf{x}_{t-1})'\mathbf{f}_t(\mathbf{r}_t) + \boldsymbol{\epsilon}_t. \tag{3.1.5}$$

The specialty of above specification is important as the output is also present as input and it helped the researchers (Gu et al., 2020) to find parsimonious nonlinear representation in their dataset.

3.2 Penalized Regressions

3.2.1 Simple Regressions: The Principles

Legendre (1805) is an early reference on least squares optimization, which makes the ideas behind linear models at least two centuries old. Given a matrix of predictors **X**, we seek to decompose the output vector **y** as linear function of the columns of **X** plus an error term $\boldsymbol{\epsilon}$. $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. The best choice of $\boldsymbol{\beta}$ is naturally the one that minimizes the error. For analytical tractability, it is the sum of squared errors that is minimized $L = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = \sum_{i=1}^{I} \epsilon_i^2$. This loss L is called the sum of squared residuals (SSR). In order to find optimal $\boldsymbol{\beta}$, it is imperative to differentiate this loss L with respect to $\boldsymbol{\beta}$ because the first order condition requires that the gradient be equal to zero:

$$\begin{aligned}\nabla_{\boldsymbol{\beta}} L &= \frac{\partial}{\partial \boldsymbol{\beta}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}}\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} \\ &= 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - 2\mathbf{X}'\mathbf{y}\end{aligned}$$

The first order condition $\nabla_{\boldsymbol{\beta}} = \mathbf{0}$ is satisfied if

$$\boldsymbol{\beta}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{3.2.1}$$

Which is known as the standard ordinary least squares (OLS) solution of the linear model.

3.2.2 Forms of penalization

Penalized regressions have been popularized since the seminal work of (Tibshirani, 1996). The idea is to impose a constraint on the coefficients of the regression so that their total magnitude be restrained. In the original paper, (Tibshirani, 1996) proposes to estimate the following LASSO model:

$$y_i = \sum_{j=1}^{J} \beta_j x_{i,j} + \epsilon_i, \quad i = 1, \dots, I, \quad \text{s.t.} \quad \sum_{j=1}^{J} |\beta_j| < \delta, \tag{3.2.2}$$

for some strictly positive constant δ. Under least square minimization, this amounts to solve the Langragian formulation:

$$\min_{\beta} \left\{ \sum_{i=1}^{I} \left(y_i - \sum_{j=1}^{J} \beta_j x_{i,j} \right)^2 + \lambda \sum_{j=1}^{J} |\beta_j| \right\}, \tag{3.2.3}$$

for some value $\lambda > 0$ which naturally depends on δ, the lower the δ, the higher the λ : the constraint is more binding. This specification is close to the ridge regression which is L^2 regularization:

$$\min_{\beta} \left\{ \sum_{i=1}^{I} \left(y_i - \sum_{j=1}^{J} \beta_j x_{i,j} \right)^2 + \lambda \sum_{j=1}^{J} \beta_j^2 \right\}, \tag{3.2.4}$$

and which is equivalent to estimating the following model:

$$y_i = \sum_{j=1}^{J} \beta_j x_{i,j} + \epsilon_i, \quad i = 1, \dots, I, \quad \text{s.t.} \quad \sum_{j=1}^{J} \beta_j^2 < \delta, \tag{3.2.5}$$

3.2.3 Predictive regressions

The introduction of penalization within predictive regressions goes back to (Rapach, Strauss & Zhou, 2013), where they are used to assess lead-lag relationships between US markets and other international stock exchanges. More recently, (Chinco, Clark-Joseph & Ye, 2019) use LASSO regressions to forecast high frequency returns based on past returns in the cross-section. (Han et al., 2019) and (Rapach, Strauss, Tu & Zhou, 2019) use LASSO and elastic net regressions to improve forecast and single out the characteristics that matter when explaining stock returns. In simple machine-learning based asset pricing, models are often built on Eq (4), but if we stick to linear relationship and add penalization terms, the model becomes:

$$r_{t+1,n} = \alpha_n + \sum_{k=1}^{K} \beta_n^k f_{t,n}^k + \epsilon_{t+1,n}, \quad \text{s.t.} \quad (1-\alpha)\sum_{j=1}^{J} |\beta_j| + \alpha \sum_{j=1}^{J} \beta_j^2 < \theta$$

where we use $f_{t,n}^k$ or $x_{t,n}^k$ interchangeably and θ is penalization intensity. Again, one of the aims of the regularization is to generate more robust estimates. If the patterns extracted hold out of sample, then the equation below will be a relatively reliable proxy of future performance.

$$\hat{r}_{t+1,n} = \hat{\alpha}_n + \sum_{k=1}^{K} \hat{\beta}_n^k f_{t,n}^k,$$

3.3 Tree-based methods

3.3.1 Simple trees: The principles

Decision trees seek to partition datasets into homogeneous clusters. Given an exogenous variable $\mathbf{Y}$ and features $\mathbf{X}$, trees iteratively split the sample into groups which are as homogeneous in as $\mathbf{Y}$ possible. The split is made according to one variable within the set of features. For technical construction of the splitting process, we follow the approach of (Hastie, Tibshirani & Friedman, 2009) that is given a sample of (y_i, x_i) of size I, a regression tree seeks the spilitting point which minimizes the total variation of the y_i inside the two child clusters. In order to do that, it first finds the best splitting point so that clusters are homegeneous in Y for each feature $x_i^{(k)}$ and then selects the feature that achieves the highest level of homegeneity. Homegeneity in trees is closely related to variance. Since we want the y_i inside each cluster to be similar, we seek to minimize their dispersion inside each cluster and then sum the two figures. We cannot sum the variances because this would not take into account the relative sizes of clusters. Hence, we work with total variation that is the variance times the number of elements in the clusters. To find the best split for each feature, we solve the $\underset{c^{(k)}}{\text{argmin}}\, V_I^{(k)}(c^{(k)})$ with

$$V_I^{(k)}(c^{(k)}) = \underbrace{\sum_{x_i^{(k)}<c^{(k)}} \left(y_i - m_I^{k,-}(c^{(k)})\right)^2}_{\text{Total dispersion of first cluster}} + \underbrace{\sum_{x_i^{(k)}>c^{(k)}} \left(y_i - m_I^{k,+}(c^{(k)})\right)^2}_{\text{Total dispersion of second cluster}}, \tag{3.3.1}$$

where

$$m_I^{k,-}(c^{(k)}) = \frac{1}{\#\{i, x_i^{(k)} < c^{(k)}\}} \sum_{\{x_i^{(k)}<c^{(k)}\}} y_i \quad \text{and}$$
$$m_I^{k,+}(c^{(k)}) = \frac{1}{\#\{i, x_i^{(k)} > c^{(k)}\}} \sum_{\{x_i^{(k)}>c^{(k)}\}} y_i$$

are the average values of Y, conditional on $X^{(k)}$ being smaller or larger than c. The cardinal function $\#\{\cdot\}$ counts the number of instances of its argument. For feature k, the optimal split $c^{k,*}$ is the one for which total dispersion over the two subgroups is the smallest. Of all the possible spilitting variables, the tree will choose the one that minimizes the total dispersion not only over all splits, but also over all variables: $k^* = \underset{k}{\text{argmin}}\, V_I^{(k)}(c^{k,*})$. After one split is performed, the procedure continues on the two newly formed clusters. Each leaf has an average value for the label, which is the predicted outcome, and this only works when the label is numerical as in our case. There are different criteria that can determine the stoppage of the splitting process. It is imperative to limit the size of the tree to avoid over fitting, the process called pruning.

3.4 Random forests

The combination or ensemble of many simple trees is random forest, which seems to be a reasonable path towards the diversification of prediction errors. The major reference for random forests is (Breiman, 2001), who proposed practical considerations for building a random forest. There are two ways to create multiple predictors from simple trees, and random forest combine both:

- The model is trained on similar yet different datasets. One way to achieve this is through bootstrapping, where the instances are resampled with or without replacement for each individual tree, yielding new training data each time a tree is built.

- The data can be altered by curtailing the number of predictors. Alternative models are built based on different set of features. The user chooses the number of features to retain and then the algorithm selects these features randomly at each try.

Hence it becomes simple to grow many trees and the ensemble is simply a weighted combination of all trees. As random forests are built on the idea of bootstrapping, they are more efficient than simple trees. (Ballings et al., 2015), (Patel et al., 2015), (Krauss et al., 2017), and (Huck, 2019) have shown their relative better performance in comparison with other algorithms. The original theoretical properties of random forests are demonstrated in Breiman (2001), where he defines the margin function as:

$$mg = M^{-1}\sum_{m=1}^{M} 1_{\{h_m(\mathbf{x})=y\}} - \max_{j\neq y}\left(M^{-1}\sum_{m=1}^{M} 1_{\{h_m(\mathbf{x})=j\}}\right),$$

where the left part is the average number of votes based on the M trees h_m for the correct class. The right part is the maximum average for any other class. The margin reflects the confidence that the aggregate forest will classify properly. Notably, (Breiman, 2001) also shows that as the number of trees grows to infinity, the accuracy converges to some finite number which explains why random forests are not prone to overfitting.

3.5 Neural network

3.5.1 Principles of a basic neural network perceptron

The emergence of neural networks dates back to the stufy of (Rosenblatt, 1958), which proposes a binary classification algorithm as a result. The nodes of the neural network are interconnected into the layers. An input layer, a hidden layer and a output layer are the typical ones that are interlinked. Learning takes place in the hidden layer, where a summation operation is performed on each before an activation function is applied on it. The signals are then passed next layers and so on. Due to the flexibility and robustness of approach, neural networks can capture non-linearities and complex interactions among variables. The model is the following:

$$f(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x}'\mathbf{w} + b > 0 \\ 0 & \text{otherwise} \end{cases}$$

The vector of weights w is a scaling parameter for the variables and the bias b determines the overall direction of the decision-making process. The value of error is computed by $\epsilon_i = y_i - 1_{\left\{\sum_{j=1}^{J} x_{i,j} w_j + w_0 > 0\right\}}$. after the initial calculations for biases b and weights w_i. The default values of the bias is set to zero s $b = w_0$ and then an initial constant is added to $x : x_{i,0} = 1$, such that $\epsilon_i = y_i - 1_{\left\{\sum_{j=0}^{J} x_{i,j} w_j > 0\right\}}$. The optimal weights in the perceptrons are only approximated, thus they do not provide a close ended solution unlike the traditional regression techniques. Initial weights can be derived by minimizing the sum of squared errors and to serve this purpose an efficient way to proceed is

- computing the model value at point where $x_i : \tilde{y}_i = 1_{\left\{\sum_{j=0}^{J} w_j x_{i,j} > 0\right\}}$.
- adjust the weight vector: $w_j \leftarrow w_j + \eta(y_i - \tilde{y}_i)x_{i,j}$.

which acts as adjusting the weight in the right direction. The learning rate often represented as scaling factor η acts similarly in both the neural networks and the decision trees. The learning will be rapid if the shifts implied by η is large, but convergence may be slow or may even not occur as a trade off. Thus, to reduce the risk of overfitting, a small η is usually preferable.

3.5.2 Multilayer perceptron (MLP)

A single perceptron can be viewed as a linear mapping of inputs and outputs to which is applied a particular activation function. This is how nonlinearity is introduced in an otherwise linear model. The idea behind neural networks is to combine multiple perceptrons together, which in essense is similar to the random forests with trees. It follows the following process:

- The data enters the network and goes through an initial linear mapping:

$$v_{i,k}^{(1)} = \mathbf{x}_i'\mathbf{w}_k^{(1)} + b_k^{(1)}, \text{for } l = 1, \quad k \in [1, U_1],$$

- A non-linear activation function f^1 transforms the data to act as an input for the next layer. With different weights and biases often calculated through back-propagation, the process of linear mapping will be repeated for each layer until the end of the network:

$$v_{i,k}^{(l)} = (\mathbf{o}_i^{(l-1)})' \mathbf{w}_k^{(l)} + b_k^{(l)}, \text{for } l \geq 2, \quad k \in [1, U_l].$$

- The layers are inter-connected such that the output of previous layer acts as an input for next layer. Yielding inputs in such a way for the next layer is essentially the same as linear mapping of inputs and outputs to which the activation function $f^{(l)}$ have been applied.

$$o_{i,k}^{(l)} = f^{(l)} \left(v_{i,k}^{(l)} \right).$$

- Finally, in the terminal stage, the outputs are aggregated from the feedback of the last layer:

$$\tilde{y}_i = f^{(L+1)} \left((\mathbf{o}_i^{(L)})' \mathbf{w}^{(L+1)} + b^{(L+1)} \right).$$

From the lens of factor investing, the characteristics of the firm initiate the process of feeding the input layer. The value of the features is multiplied by the initial weights and a bias is added, in the first step. The process is performed on all the units of first layer, such that the output of first step, which is a linear combination of the input is then transformed by the activation function. All of the values from the first layer are combined and then fed to the second layer following the same process. These iterations are performed until the end of network. The objective of last layer is to produce a linear output that may correspond to a categorical label or numerical value. The output is a single number in case of numerical value and a vector equal to the number of categories if it is categorical label. The outperformance of neural networks can be attributed to their universal approximations. Meaning that a simple network with few layers can approximate the underlying function with arbitrary precision given a bounded continuous function. (Cybenko, 1989). For more early references, (Du & Swamy, 2013), (Goodfellow, Bengio & Courville, 2016) and (Guliyev & Ismailov, 2018) for recent results.

We can now formally define a single-layer perceptron as:

$$f_n(\mathbf{x}) = \sum_{l=1}^{n} c_l \phi(\mathbf{x}\mathbf{w}_l + \mathbf{b}_l) + c_0,$$

where ϕ is a non-constant bounded continuous function.
The training of neural networks is done by minimizing error function subject to some penalization just like for other machine-learning tools:

$$O = \sum_{i=1}^{I} \text{loss}(y_i, \tilde{y}_i) + \text{penalization},$$

where y_i are the true values of all instances and $\tilde{y}_i$ are the values obtained by the model. The weights and biases of all the units of all layers are adjusted in the training of neural network, such that O defined above is the smallest possible, which is performed via gradient descent with following equation:

$$\mathbf{W} \leftarrow \mathbf{W} - \eta \frac{\partial D(\tilde{y}_i)}{\partial \mathbf{W}}. \tag{3.5.1}$$

The ultimate goal of the decision trees is to create homogeneous clusters as explained preiously. In contrast, the nerural network works to reduce the error in prediction $\tilde{\mathbf{y}}_i$ and a target label y_i.. The choice of suitable activation function is more important for a classification problem specially at the end of the network.

3.6 Evaluation measures

The measures we will use to evaluate the machine learning techniques are mean squared error (MSE) and hit ratio. The MSE measures the variance between the training and test sample i.e. the average squared difference between the estimated value based on training sample and actual value based on test sample. However, the MSE is is harder to interpret for an average investor because it is complicated to map it into intuitive financial indicator. The hit ratio is more ogranic in this perspective as it demonstrates the proportion of times the prediction guesses the return correcly i.e. if we take a long position based on positive signal and short position based on negative ones, the hit ratio indicates the proportion of correctly identified expected direction.

3.7 Data sources and preprocessing

We have resided to utilize two different datasets in this thesis for the comparison purposes. In the process of building the ensembles, the machine learning techniques that we are employing are highly correlated in terms of extracting the variables especially in the case of US. As most the models are essentially extracting the same pattern, there is very less diversification benefit that is available if an ensemble has to be created. It makes it very difficult to even beat a simple equally weighted ensemble. To draw a comparison and evaluate the feasibility of ensembles we will also train our models in the data of Pakistan in addition to US. The data for US stocks is taken from (Coqueret & Guida, 2020a) already pre-processed, while we manually extracted the required variables from publicly available banks traded on Pakistan Stock Exchange (PSX). The financial reports are available at quarterly frequency from the PSX website (PSX, 2021). Monthly price and volume information is extracted from the company archives of business recorder website (Recorder, 2021). We have choosen the time-period from 12/31/1999 to 01/01/2019 for 1200 US stocks with a total of 93 variables and 9/30/2005 to 9/30/2020 of 19 major banks in Pakistan with a total of 22 variables. A full list is provided in the appendix. After the data extraction, we join the monthly price and volume information with quarterly company financials by replacing the missing points with the closing values. We utilized the exploratory data analysis of the features to reveal the distributions of data. As an example, the distribution of the market capitalization before uniformization is shown in figure 3.1. This plot clearly shows the positive skewness in the market capitalization, which is not ideal to the analysis we are performing in the study. To circumvent the issue we have to resort to some normalization technique, as applied in figure 3.3 before inputting the data to the model.

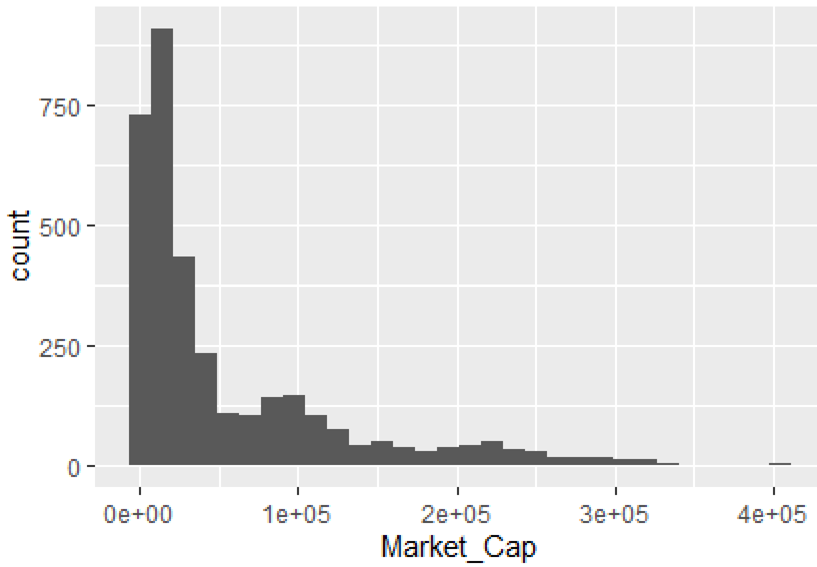

Figure 3.1: Distribution of market capitalization before uniformization Pakistan

We further show a box plot illustrating the distribution of correlations between the selected features and the returns in figure 3.2. The correlations are computed over the whole cross-section of stocks. They are mostly located close to zero, but there are periods which experience exteme shifts as demonstrated by the black circles representing outliers. Further, price to sales ratio, the price to book ratio, and the relative strength indicator are the predictors with positive median correlation with returns.

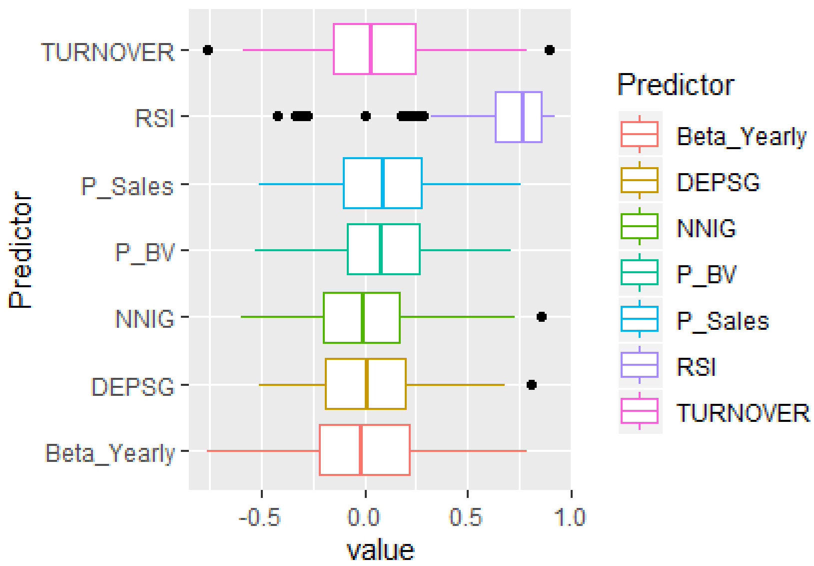

Figure 3.2: Distribution of correlation of factors with returns Pakistan

As demonstrated above, the distribution of some variables is not normal. Additionally they also differ in the measurement scale, i.e. market capital is measured in billions, while returns does not have any units. Both of these concerns render any type of machine learning analysis useless. We thus need to homogenize the data as most models like neural networks perform much better when predictors have similar scales and normally distributed. The convention is to scale inputs so that they range in [0,1] before sending them through the training of neural networks (Freyberger et al., 2020). The simplest way to uniformization is to rank the firms according to some indicator and assign the values between 0 to 1. It is done such that at each date the variables are processed so that the firm with the smallest indicator is equal to 0 and 1 for the largest firm. After this step, all of the features are uniformly scaled accross the cross-section of assets and are comparable in magnitude. Mathematically, if we write x_i for the raw input and $\tilde{x}_i$ for transformed data, common scaling methods include:

- standardization: $\tilde{x}_i = (x_i - m_x)/\sigma_x$.
- min-max rescaling: $\tilde{x}_i = (x_i - \min(\mathbf{x}))/(\max(\mathbf{x}) - \min(\mathbf{x}))$
- uniformization: $\tilde{x}_i = F_{\mathbf{x}}(x_i)$, where $F_{\mathbf{x}}$ is the empirical c.d.f of x. In this case the vector $\tilde{x}$ is defined to follow a uniform distribution over [0,1].

We have first taken the log of the indicators with bigger values before applying the 0 to 1 uniformization to data. We now present, as an example the market capitalization in figure 3.3 showing the smoothness achieved after the pre-processing.

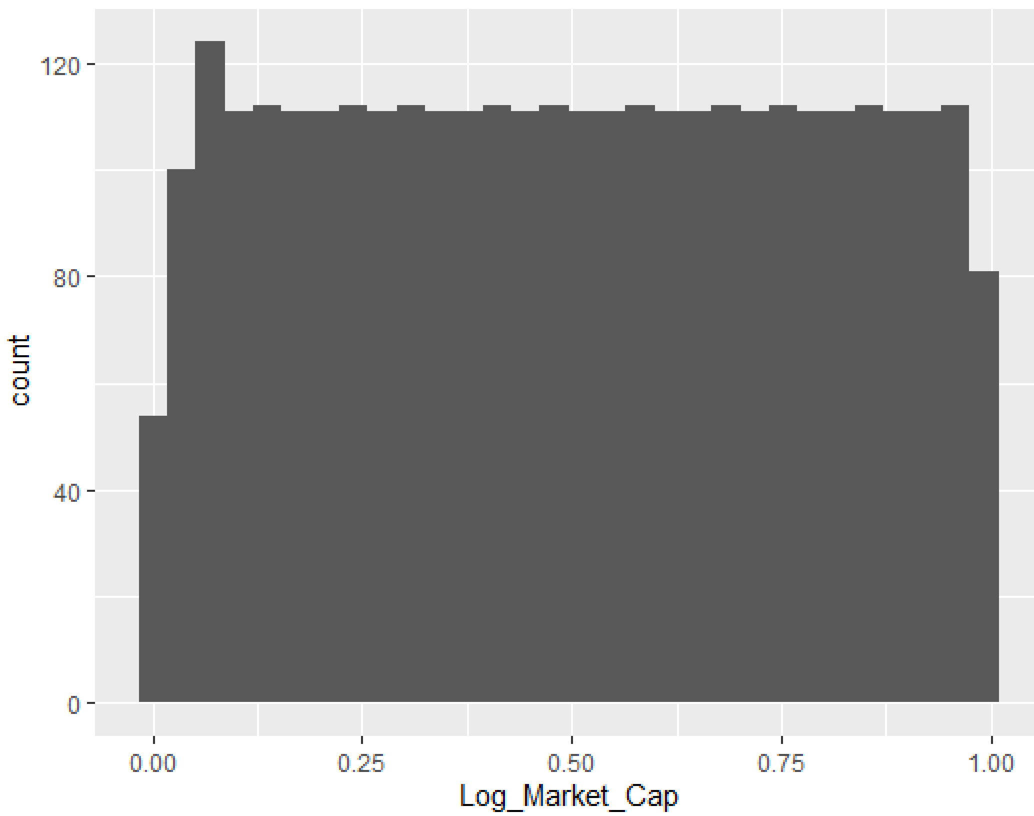

Figure 3.3: Distribution of market capitalization after uniformization Pakistan

Chapter 4

Results and Economic Interpretation

4.1 Illustration and results of Penalized Regression

The outcome is different for every penalization technique, thus justifies a separate treatment. Mechanically, as λ increases, all of the coefficients of the ridge regression slowly decrease in magnitude towards zero. The convergence is much smoother as shown in the figure 4.1. Clearly, the ridge regression gleans market capitalization as the most dominating predictor with large negative coefficients. The negative coefficients indicate the presence of size anomaly in our sample, according to which small firms experience higher future returns compared to their large counterparts. These results are in line with (Astakhov et al., 2019) and (C. Asness et al., 2018). Furthermore, our sample is also confirming the contribution of momentum, price to book ratio, price to earning ratio, and net profit margin etc. in explaining significant portion of variation in stock returns. Other small contributors include earning per share, free cash flow yield, earning before interest and tax over operating profits, and interest expense etc.

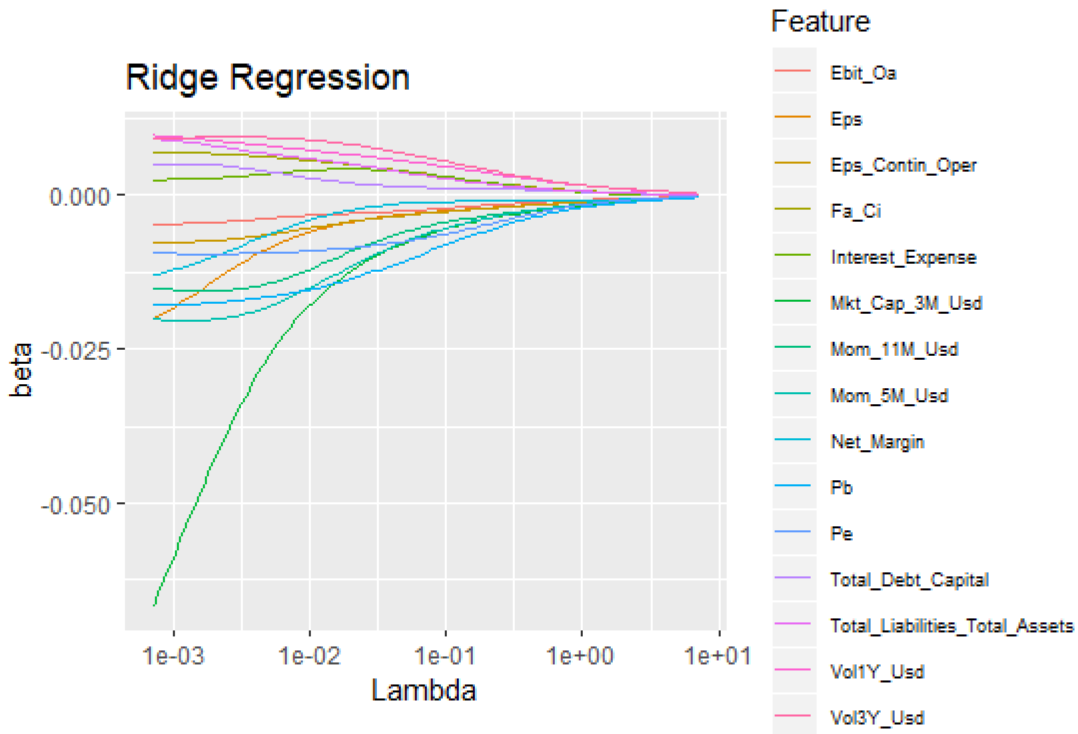

Figure 4.1: Ridge Regression US Stocks

The figure represents the LASSO model, where the convergence is brutal as some coefficients shrink to zero very quickly. For instance, the convergence of earning before interest and tax over operating profit to zero is quick. Other variables that persist the penalization longer i.e. for λ sufficiently large, the market capitalization resisted the most unlike the ridge regression where zero value is only reached asymptotically for all coefficients. Again the negative sign with the coefficient of market capitalization signifies the presence of size anomaly in the US data. Similar to ridge regression the moderate and small contributors remain the same in LASSO model.

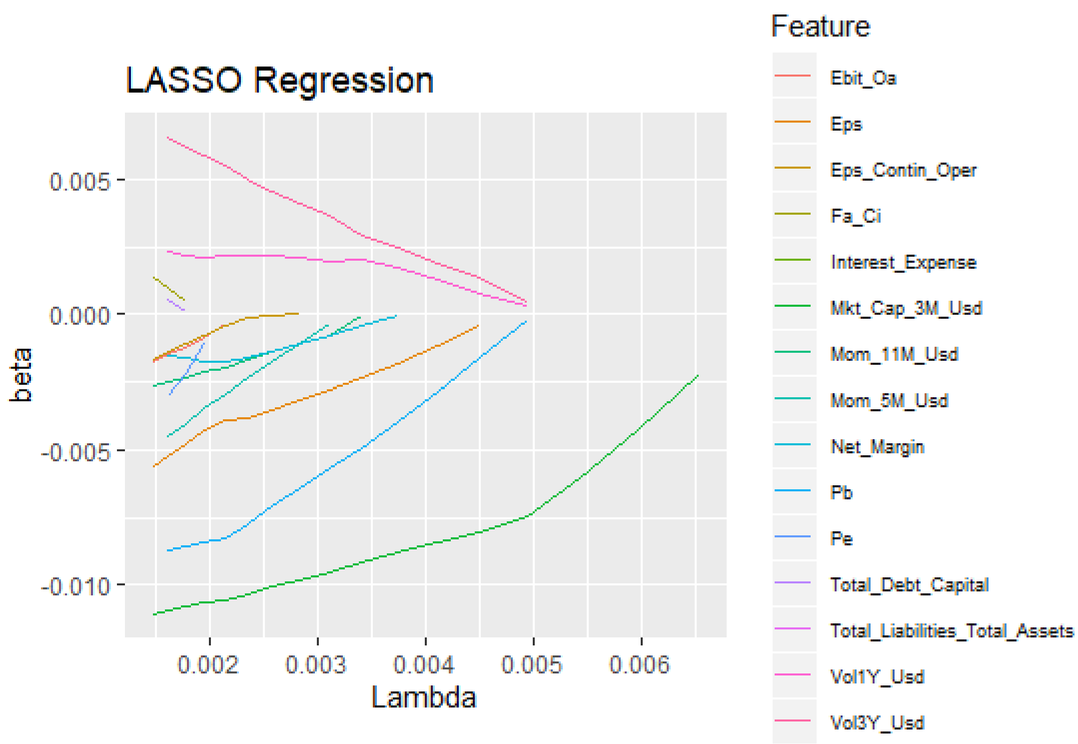

Figure 4.2: LASSO Model US Stocks

Elastic nets, by definition are the combination of both ridge and LASSO models, meaning that as long as the value of alpha is greater than 0, the shrinkage quality of LASSO will be preserved. So, LASSO and elastic net functions to shrink the dimensionality of variable space and hence an ideal candidate model selection tool. The figure 4.3 shows the variable importance of the elastic net approach, which is very much in line with the results of ridge and LASSO models.

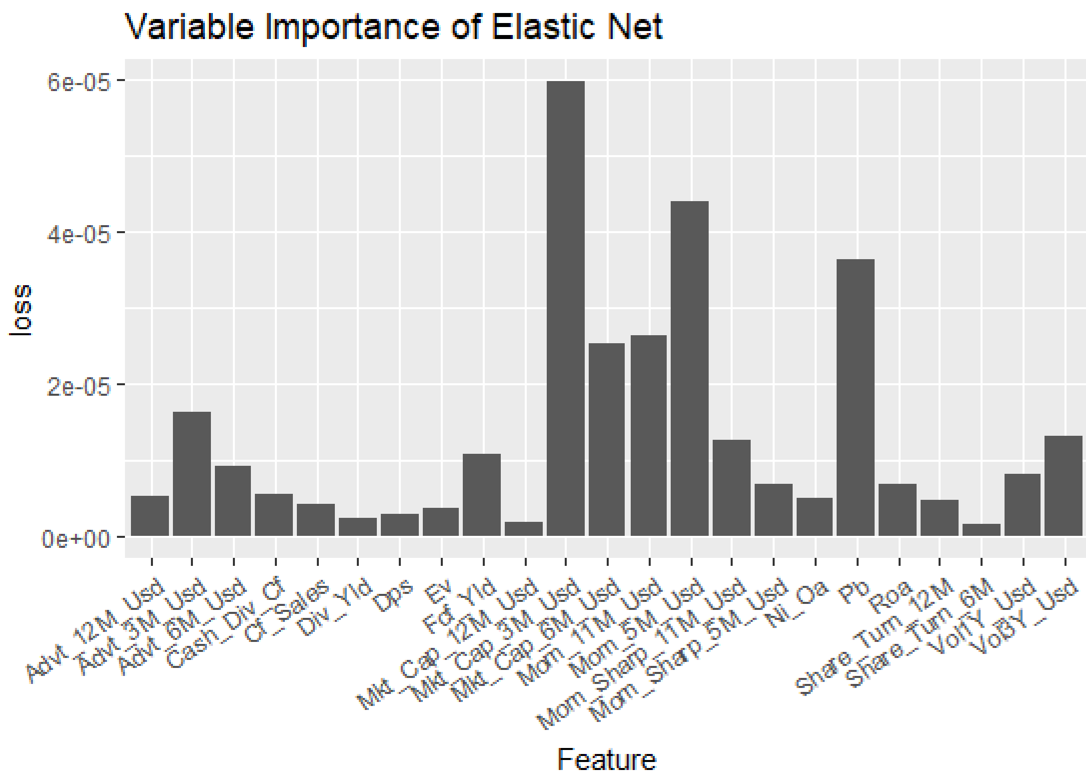

Figure 4.3: Variable Importance of elastic net

From the table 4.1, elastic net outperforms both the LASSO and ridge regressions in both the terms of our evaluation measures, the mean squared error (MSE) being the lowest at 3.6% and the highest hit ratio of 54.60%. The MSE measures the variance between the training and test samples, while the hit ratio is the proportion of times that the prediction guesses the sign of the return correctly in the testing sample. A natural benchmark for hit ratio is 50% but to account for the transaction costs, adding 1% on average will be more prudent. All of our penalized regression models were able to beat the benchmark considering transaction costs by reasonable margin.

Method	MSE	Hit Ratio
LASSO	0.03703482	0.5340689
Ridge	0.03701371	0.5363682
Elastic	0.03699696	0.5460346

Table 4.1: Mean squared error and Hit ratio of penalized regressions

excel

4.2 Illustrations and results of Decision trees

The Rpart package is used for implementation of decision trees and to reduce the risk of overfitting, several criteria to stop the splitting process of decision trees are presented below:

- Impose a minimum number of instances for each leaf ($minbucket$ function in Rpart), which ensures that each final cluster is composed of a sufficient number of observations. The minimum number of observations required in each terminal nodes are set to 3500.
- Similarly, it can also be imposed that the cluster has a minimal size ($minsplit$ function in Rpart) before considering any further split. The minimum number of observations required to continue splitting are set to 8000.
- Require a certain threshold of improvement in the fit (cp function in Rpart) such that if a split does not sufficiently reduce the loss, it can be deemed unnecessary. The precision measure is set to be 0.0001.
- Limit the depth of the tree ($maxdepth$ function in Rpart), which is defined as the overall maximum number of splits between the root and any leaf of the tree. The maximum depth is set to be 3 in this study.

The figure 4.4 represents a characteristics based decision tree where the dependent variable is the 1 month future return. The convention in the representation of trees is at each node, a condition describes the split with a true or a false expression. If the statement is true, the observation goes to the left cluster and if not, it goes to the right cluster. Given the full sample, initial split is performed according to the market capital. If the market capital is greater than 0.15, the instance is placed on the left bucket, otherwise it goes to the right bucket. Similarily, in the next left node if the price to book ratio fulfills the benchmark of 0.025, it gets split further until maximum depth is reached. The final returns at the bottom are color coded from left to right, left being the lowwest at 0.39% and right the highest at 11%. At the top of the tree with all instances one month future return is 1.4%.

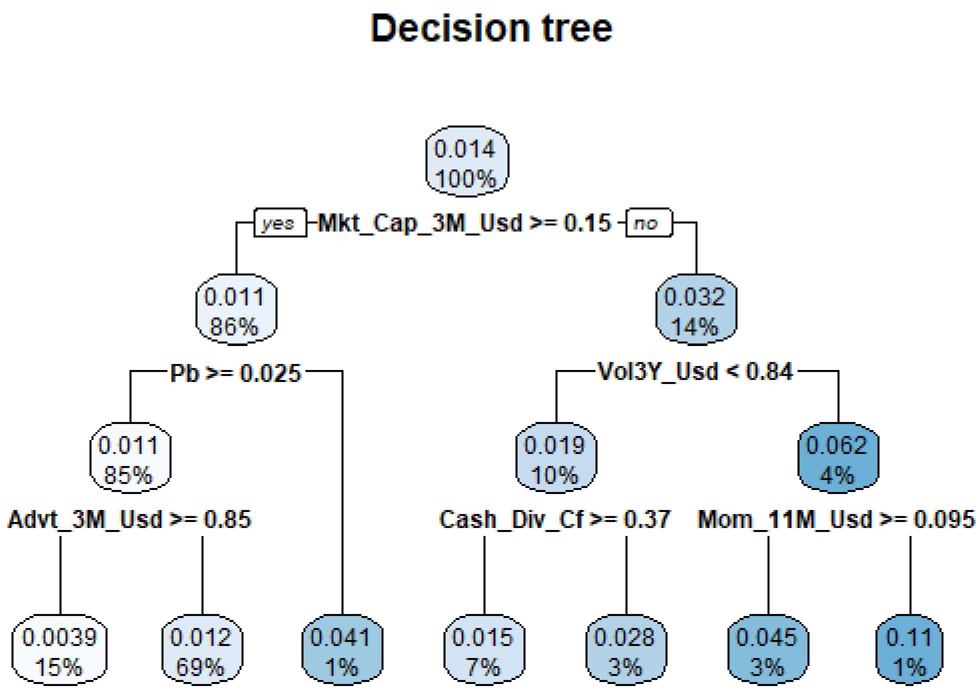

Figure 4.4: Graphical representation of decision trees

The figure 4.5 represents the feature importance of the decision trees, where market capitalization, momentum, and volatility are the most significant ones. The results in terms of varibale selection are quite similar to penalized regressions. The prevalence of size and momentum anomalies are reiterated, while volatility measure also becomes significant.

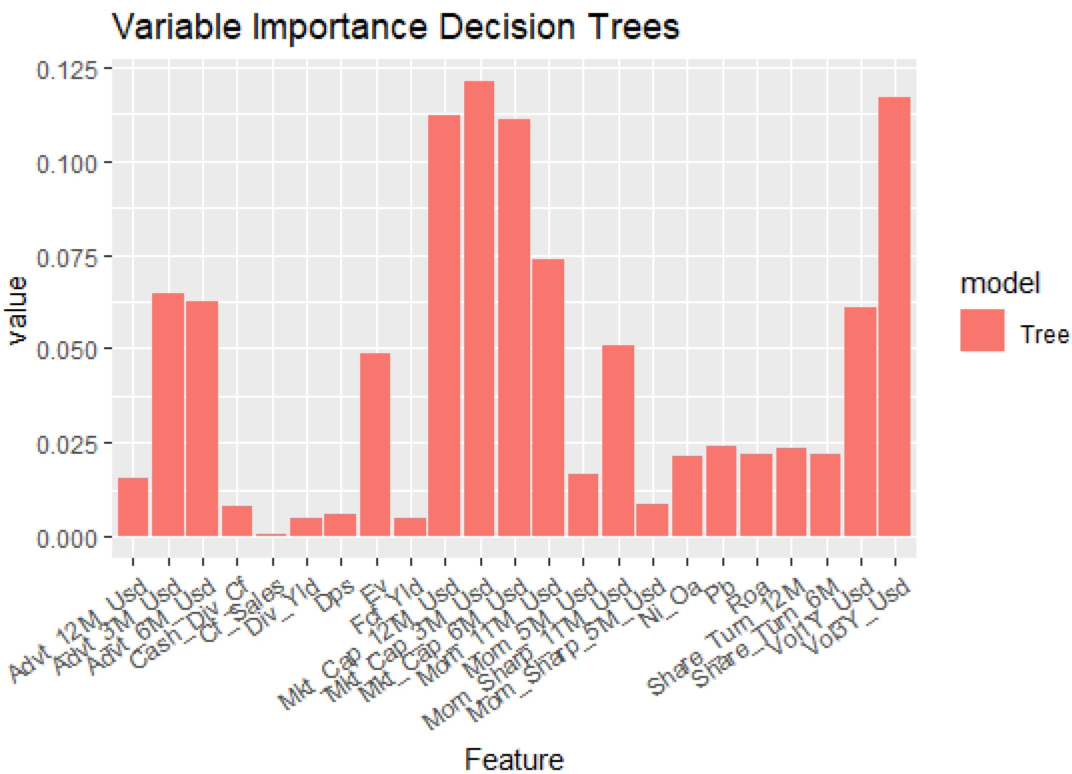

Figure 4.5: Variable importance of Decision trees

From the table 4.2 below, the hit ratio of decision tree is 54.6% which is exactly the same as the hit ratio of elastic net. However, the decision trees slightly improved upon in reduction of mean squared error in comparison to penalized regressions.

Method	MSE	Hit Ratio
Decision Tree	0.0369881	0.5460346

Table 4.2: MSE and Hit ratio of decision trees

4.3 Illustrations and Results of Random Forest

The decision trees are known to over fit the training sample, although we have put together several controls to avoid it by limiting the size of trees. Random forests are more efficient than simple regression trees as they are built on the idea of bootstrapping. Several recent studies (Krauss et al., 2017) and (Huck, 2019) have shown their successful implementation of random forests, although the original theoretical properties of random forest are demonstrated in (Breiman, 2001). Sequentially, we first make a comparison of the variables extracted by decision trees and random forests in figure 4.6. We can observe from the figure 4.6 that many of the idiosyncrasies of decision trees got tamed. Specially the impact of market capitalization reduces in magnitude in random forests unlike both the decision trees and penalized regressions. There is an increase in magnitude in both the momentum anomaly and price to book ratio, while volatility has mixed results. Other significant contributors include cash flow to sales and dividend yield ratios, which are very unique to this study.

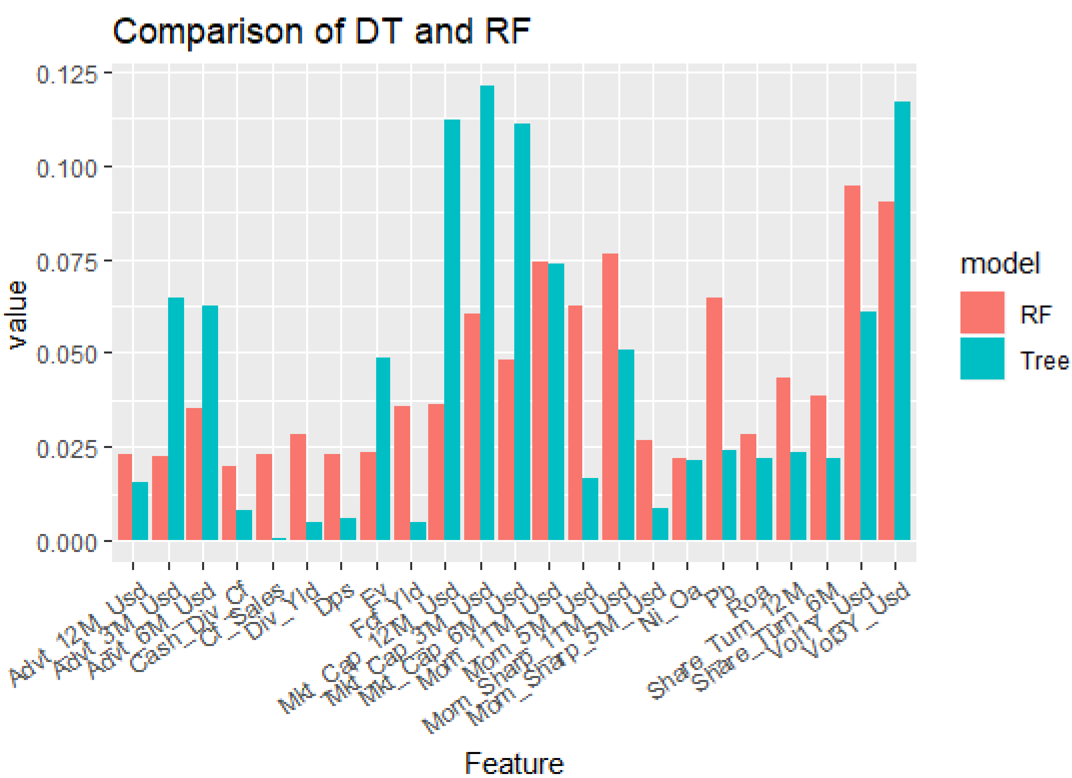

Figure 4.6: Comparison of Decision trees and Random forest

Finally we present the table 4.3 stating mean squared error and hit ratio of random forests. Random forest has certainly reduced the over fitting problem of the trees as demonstrated by the improvement in MSE with a negligible reduction in hit ratio.

Method	MSE	Hit Ratio
Random Forest	0.03689782	0.5434993

Table 4.3: MSE and Hit ratio of Random forest

4.4 Illustration and results of Neural Networks

It is important for the neural networks to have the number of parameters smaller than the number of instances. It is preferable to have a large sample, which is readily available for US stocks from many sources. We start the neural network by calculating the number of parameters that includes the weights and the biases to be estimated in a network.

- In the first layer, we have $(U_0 + 1)U_1$ parameters, where U_o are the number of independent variables represented as columns in $\mathbb{X}$ and U_1 depicts the number of neuron in the first layer.
- The number of parameters for hidden layers are $(U_{l-1} + 1)U_l$..
- Finally, $U_L + 1$ are the number of parameters for the output layer.
- The total number of the parameters that require optimization are

$$\mathcal{N} = \left(\sum_{l=1}^{L} (U_{l-1} + 1)U_l \right) + U_L + 1$$

The table 4.7 shows the number of parameters in each layer of our neural network. It is a forward pass list meaning the order is from input to output. The number of parameters for the first layer with 24 neurons is 93 the number of features plus one for the bias multiplied by 24, which makes it 2256 and similarly so on.

Method	Layer name	Number of neurons	Number of parameters
Neural network	Input layer	24	2256
	first layer	16	400
	Hidden layer	8	136
	Output layer	1	9
Total trainable parameters			2801

Table 4.4: Architectural choice of neural network

The improvement in the loss function as the number of training epochs increases is shown in the figure 4.7 below. The learning is quite rapid at the beginning of training process but converges suddenly to a point where no additional epoch offers any improvement.

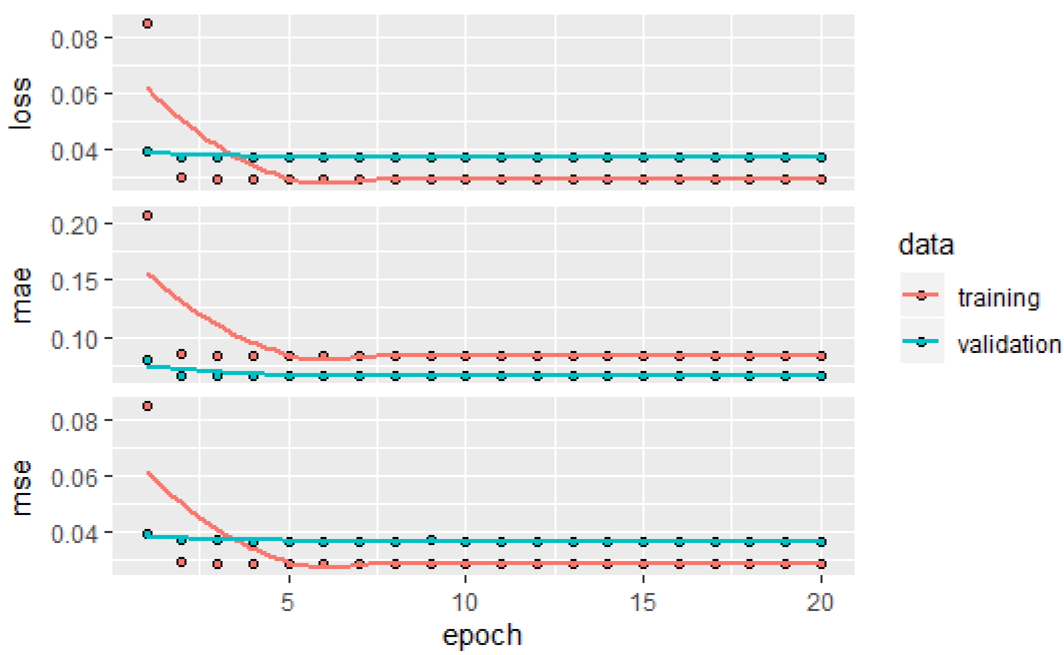

Figure 4.7: Neural network training

The figure 4.8 shows the variable importance of the neural network. Interestingly unlike other algorithms the most highly rated variable under neural network is the average daily volume over 3 months. Other unique findings of our neural network is the prevalence of capital expenditure to price to sales cash flow ratio, which may require further investigation of other asset classes to confirm the finding. Other relevant factors include market capitalization, momentum, dividend yield etc. which are almost same across all the techniques we have employed.

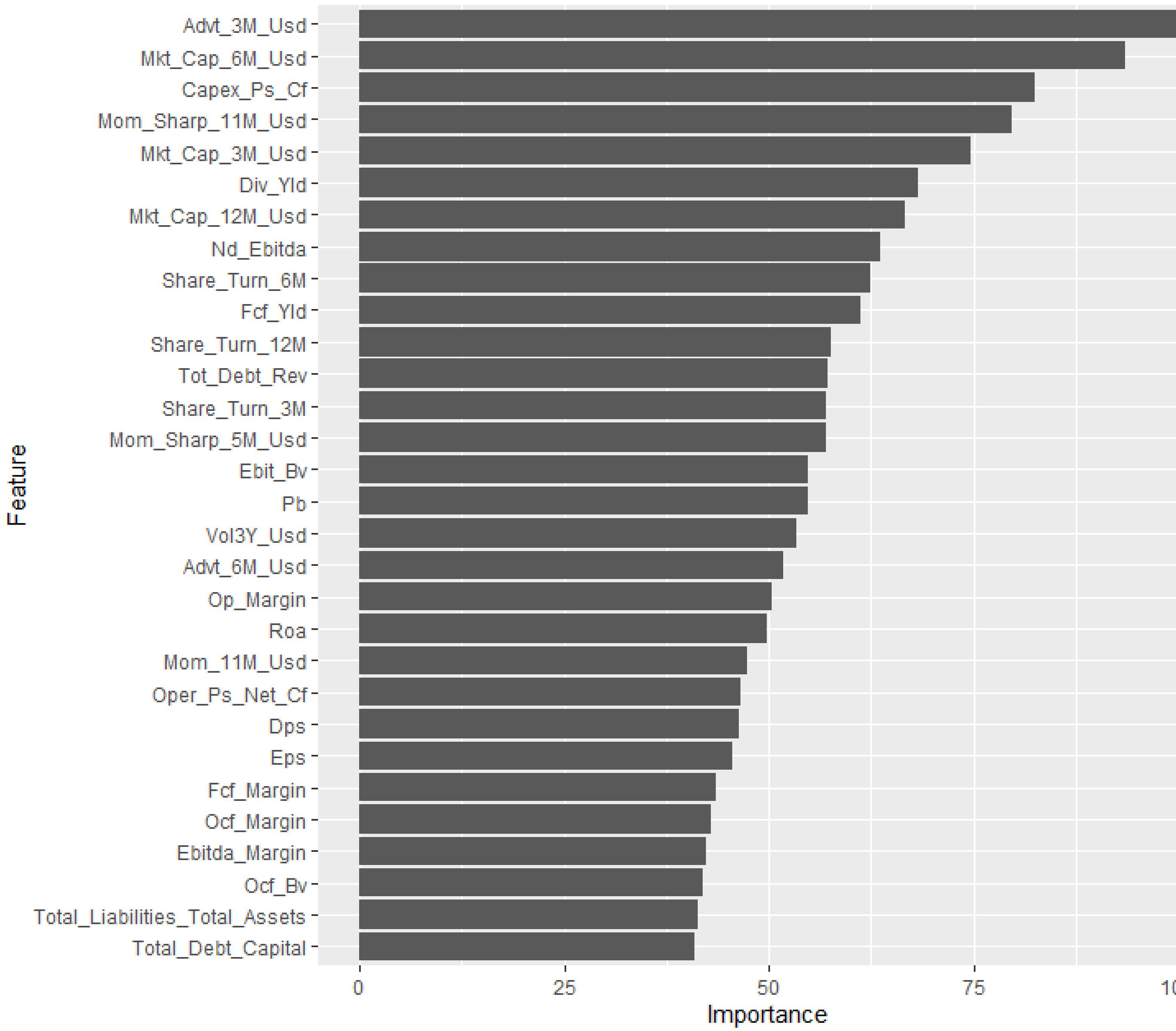

Figure 4.8: Neural network features

Finally we present the accuracy measures of our neural network in table 4.5. The hit ratio for all models is very competitive, but by far neural networks performed best in reducing the mean squared error in comparison to other models.

Method	MSE	Hit Ratio	Loss	MAE	Test Loss
Neural Network	0.02944842	0.5460346	0.02905057	0.08345813	0.03698535

Table 4.5: Accuracy measures of Neural network

Chapter 5

Ensembles

5.1 Illustration and results of Ensemble Models

In order to build an ensemble, we gathered the predictions and corresponding errors of all the models we have constructed above. A correlation matrix is then computed and shown in table 5.1. From the table 5.1, all of the models are strongly correlated to each other meaning that most of the indicators that are choosen by the models are grossly similar. There are little diversification benefits that are available in the ensemble as the models fails to generate heterogeneity in their predictions. Correlation Matrix of models in Training sample:

	Pen_reg	Tree	RF	NN
Pen_reg	1.0000000	0.9982507	0.9977416	0.9982573
Tree	0.9982507	1.0000000	0.9981544	0.9984143
RF	0.9977416	0.9981544	1.0000000	0.9987878
NN	0.9982573	0.9984143	0.9987878	1.0000000

Table 5.1: Correlation matrix in training sample for US

The table 5.2 compares the training accuracy of models by computing Mean Absolute Error (MAE) in Training sample. The regression trees performed the worst and the best performing machine learning engine in the training sample is random forest.

Pen_reg	Tree	RF	NN
0.08345916	0.08366795	0.08342645	0.08357141

Table 5.2: MAE of models in training sample

Now we compare the accuracy of our models in the testing sample, again the correlations 5.3 amongst the different techniques are high but interestingly penalized regression generalizes the training from training sample in test sample the best as shown in table 5.4, meaning that elastic net which we have used as a representative of penalized regression have the best out of sample performance.

	Pen_reg	Tree	RF	NN
Pen_reg	1.0000000	0.9985518	0.9979074	0.9979575
Tree	0.9985518	1.0000000	0.9985593	0.9986113
RF	0.9979074	0.9985593	1.0000000	0.9989766
NN	0.9979575	0.9986113	0.9989766	1.0000000

Table 5.3: Correlation matrix in testing sample of US

Pen_reg	Tree	RF	NN
0.06618181	0.06650492	0.06684155	0.06682563

Table 5.4: MAE in test sample of US

Below we calculate the optimized weights of models 5.5 based on the training data where $\mathbf{w}^* = \frac{(\mathbf{E}'\mathbf{E})^{-1}\mathbf{1}_M}{(\mathbf{1}'_M\mathbf{E}'\mathbf{E})^{-1}\mathbf{1}_M}$ and the corresponding MAE 5.6 of optimized model is shown.

Pen_reg	-0.54785327
Tree	0.16153549
RF	1.44102545
NN	-0.05470766

Table 5.5: Weights of optimized ensemble US

```
0.06763758
```

Table 5.6: MAE of optimized ensemble

Since the correlation between the models is high and thus the diversification benefits are very limited, we consider an equally weighted ensemble performance 5.7, which outperforms the optimized ensemble as shown below:

```
0.06627278
```

Table 5.7: MAE of equally weighted ensemble

5.1.1 Inclusion of macro-economic indicators

Finally, we extend our analysis to include macro-economic indicators. We included term spread, CPI, GDP, and unemployment rate. We create a decision tree 5.1 that tries to explain the accuracy of models as a function of macro-variables. One big cluster represents 92% of predictions. It corresponds to the periods when the term spread is above 0.29. The other two groups are determined according to the level of GDP. If the latter is above from the index value of 100, the average of absolute error is 6.9%. The last number of 12% indicates that when the term spread is low and GDP negative, the predictions of the models are not trust worthy because the errors have a magnitude that is twice as large as in other periods. So an important deduction is not to use ML-based forecasts in volatile economic conditions.

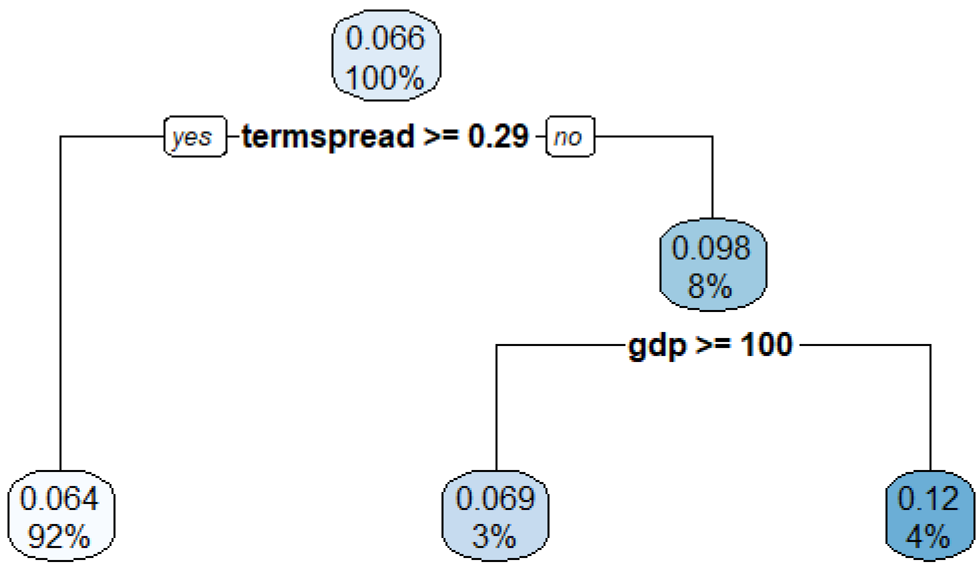

Figure 5.1: Decision tree of MAE appropriation of equal ensemble to economic indicators

5.2 Machine learning on the data of the Banks of Pakistan

5.2.1 A short glimpse into the feature selection

Since the focus of this section is to study the feasibility of ensembles on a different dataset, we therefore provide only a short glimpse of features selected accross the different techniques. The figures 5.2, 5.3, and 5.4 presents the variables selected by elastic net, tree based models, and neural networks respectively. The common indicator that is most prevalent in all of the techniques is the relative strength indicator (RSI) that reflects the momentum effect first documented by (Jegadeesh & Titman, 1993) in the prices of stock. RSI developed by (Wilder, 1978) measures the magnitude of recent prices to evaluate an over-valued or undervalued conditions in the price of a stock. Other researchers including (Rasheed, Saood & Alam, 2019) and (Tauseef & Nishat, 2016) also found similar prevalence of momentum effects in their respective samples of Pakistan stock exchange market. The beta and the size anomaly as shown in figure 5.2 are also in line with results of (Iqbal & Brooks, 2008) and (Jan, 2019) respectively. It is important to note that both the market capital to revenues ratio and price to sales ratio are effectively measuring the same metric. We have not found any study in the literature especially in the case of Pakistan, which documented price to sales ratio significantly explaining cross-section of equity returns. Other characteristics that are unique to this study and plays an important role in determining the future returns of Banks especially in our sample of Pakistan are shown in the figures below:

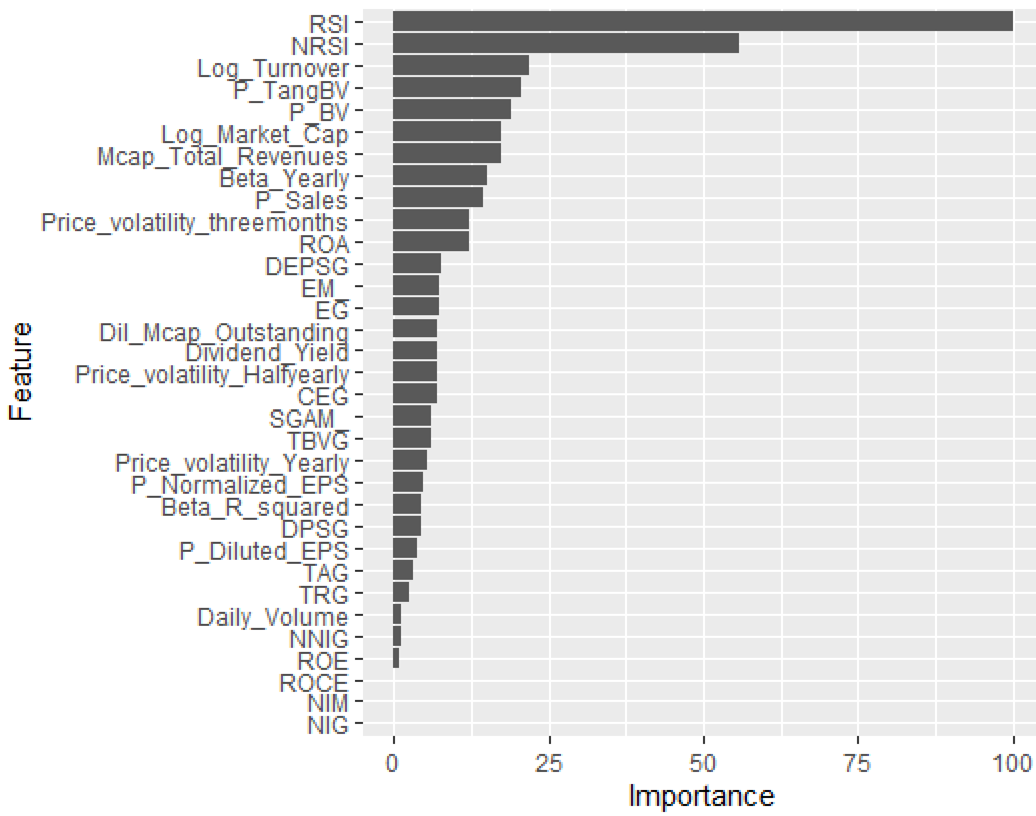

Figure 5.2: Feature selection elastic net Pakistan

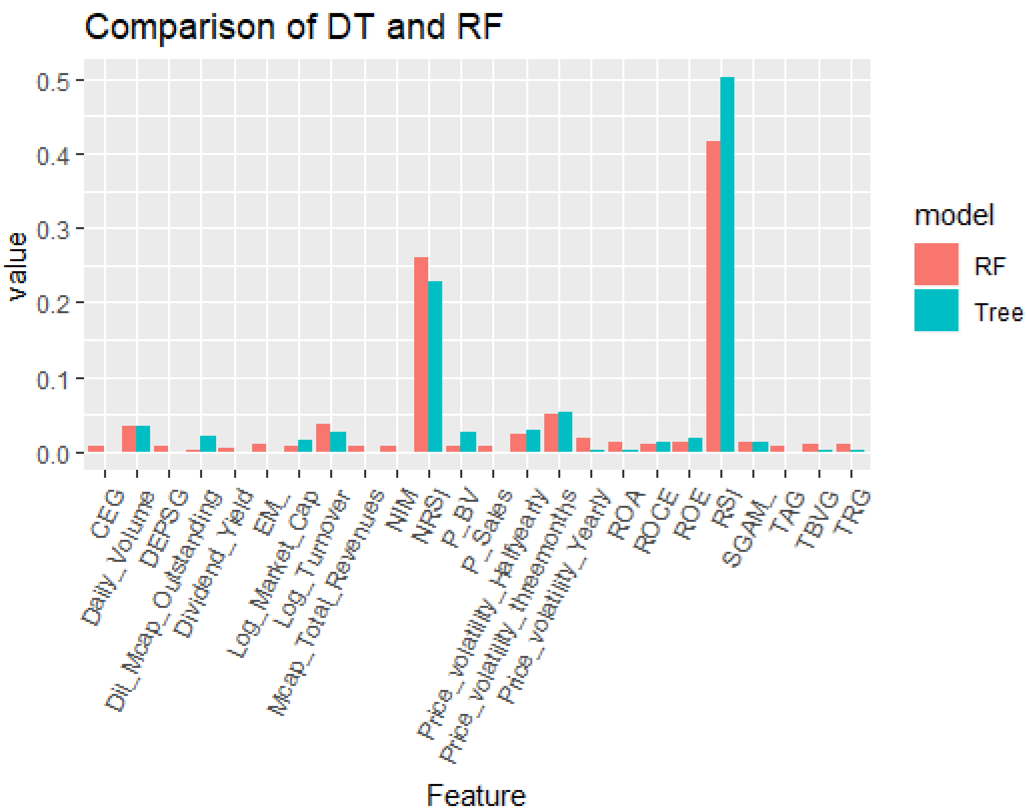

Figure 5.3: Feature selection comparison of trees and random forest Pakistan

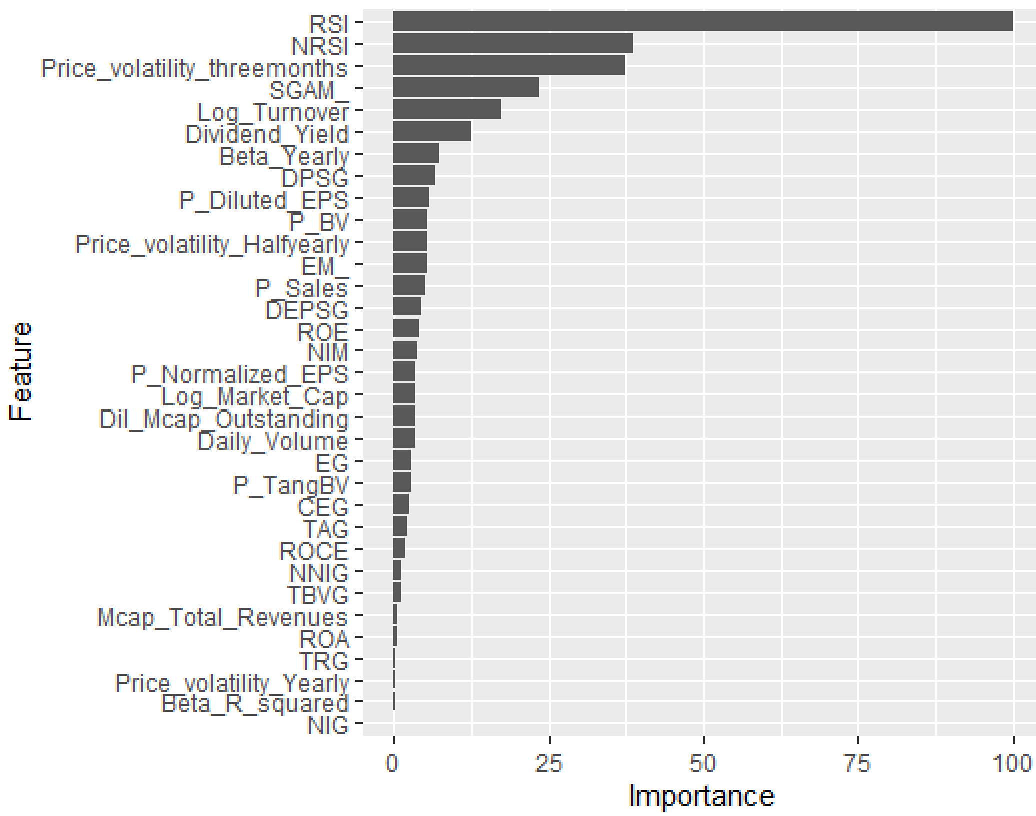

Figure 5.4: Feature selection Neural network Pakistan

Method	MSE	Hit Ratio
Pen reg	0.003749399	0.8232469
Trees	0.004360869	0.8088377
Random Forest	0.003647283	0.8386167
Neural nets	0.009174978	0.5014409

Table 5.8: MSE and Hit ratio of all methods

5.2.2 Ensemble methods on Pakistan

In order to build an ensemble, we gathered the predictions and corresponding errors into a matrix. A correlation matrix is then calculated 5.9, the models have a reasonable success to generate heterogeneity in their predictions. It essentially means that the variables extracted by the models are diversified enough but that could be due to a relatively small sample size of Pakistan. This is an interesting direction for future research and we will further discuss it in the final section below.

	Pen	Tree	RF	NN
Pen	1.0000000	0.8297448	0.9142431	0.7027043
Tree	0.8297448	1.0000000	0.8156952	0.6511078
RF	0.9142431	0.8156952	1.0000000	0.9030927
NN	0.7027043	0.6511078	0.9030927	1.0000000

Table 5.9: Correlation matrix in training sample of Pakistan

The figure 5.10 shows the mean absolute error of the tools and based on this measure random forests is the best performing technique in our testing sample. Penalized regressions also performed relatively well compared with simple trees and neural networks on the training data : MAE of ensemble on Pakistani Data:

Pen	Tree	RF	NN
0.05582614	0.05702258	0.05092703	0.08472196

Table 5.10: MAE of ensemble on Pakistan training sample

Below we calculate the optimized weights of models 5.11 based on the test data and the corresponding MAE of optimized model 5.12 is shown. Our optimized ensemble is performing almost as good as the equally weighted ensemble 5.13, but future improvements are still required. Weights in optimized ensemble:

```
Pen   -0.6089056
Tree   0.3161692
RF     2.0586458
NN    -0.7659094
```

Table 5.11: Weights in optimized ensemble of Pakistan

```
mean abs (optimal unconstrained test))
0.046219
```

Table 5.12: MAE of optimized ensemble of Pakistan

```
mean abs(Equally weighted test))
0.04197303
```

Table 5.13: MAE of equal weighted ensemble of Pakistan

We gained significant improvements in terms of the diversification benefits of our ensemble. Now we show the correlation matrix in the testing sample 5.14, MAE on testing sample 5.15, and a comparison of the performance between optimal combination and equally weighted ensembles. The random forests have the best out-of-sample performance in comparison to the penalized regression, decision trees, and neural networks respectively in the sample of Pakistan.

	Pen_reg	Tree	RF	NN
Pen_reg	1.0000000	0.7177480	0.8321413	0.4428607
Tree	0.7177480	1.0000000	0.7008418	0.4290073
RF	0.8321413	0.7008418	1.0000000	0.8219463
NN	0.4428607	0.4290073	0.8219463	1.0000000

Table 5.14: Correlation matrix in testing sample of Pakistan

Pen_reg	Tree	RF	NN
0.04624148	0.04905040	0.04077325	0.06865359

Table 5.15: MAE on testing sample of Pakistan

Chapter 6

Conclusion and Recommendations

6.1 Outcome of the study

We have analyzed and compared several machine learning techniques namely penalized regressions, trees-based methods and neural networks in the markets of US and Pakistan. We have also tested the formation of ensembles and tested them in both datasets. The results are mixed; our optimal combination of ensemble was not able to beat the equally weighted benchmark in the US market, while the performance is enhanced when the ensembles were made on the data from Pakistani Banks, but still the optimal ensemble does not comprehensively beat the equally weighted ensemble. Thus, an important future direction is to work on some combination of techniques that yield improvements over simple equally weighted ensembles.

We interpreted the traditional black-boxes and shown the variable importance in each network. The most significant variable in the US market is the market capitalization that confirms the presence of size anomaly. Other significant variables in the sample of US includes momentum effect, volatility of stocks, and price to book ratio among others. Most significant outcome is that both of the technical and fundamental indicators are included in significant predictors of US stocks. Furthermore, in general our results are promising with the hit ratio of neural networks, decision trees and elastic net around 54.6%, which is well above the benchmark of 52% considering transaction costs in the US market. Lastly for the US market 5.1, inclusion of macro-economic variables reveal that our predictions are reliable in normal economic conditions and machine learning tool should be avoided in highly volatile markets as in recession.

Although the original purpose of utilizing the data of Pakistan was to evaluate the performance of the optimal ensemble as stated above. But in the process of building the ensembles, we also tested the MSE and hit ratio of in the testing sample. We achieved considerable enhancements in terms of both the measures with the hit ratio averaging more than 80% and MSE of 0.4%. The most prevalent variables in the banks of Pakistan include the momentum indicator, price to book value, the shares turnover, market cap, and price to sales ratio among others. Machine-learning techniques have significantly more value in the emerging markets as the markets are more informationally inefficient in comparison to the developed markets like US. However, it requires more research to cover other developing markets to see if these results persist. It is therefore a very compelling research idea for future researchers to include more firm specific characteristics in their analysis.

6.2 Challenges and Future Directions

The application of machine learning techniques poses many significant challenges for financial forecasting that includes modeling of highly complex, multidimensional, and noisy data series. As already discussed above, the simple linear ensembles do not beat the equally weighted ensembles in general, but an investigation into the application of stacked ensembles for capturing trading patterns in both the supervised and un-supervised learning presents an important future direction in the field of financial forecasting. As we have not found much literature on this topic, so as a policy recommendation, we suggest to look forward for utilizing the brute force of machine-learning techniques to innovate on mutiple-staged ensembles.

Along many other challenges, the core one remains the feature engineering in a high dimensional space with minimum loss of valuable information. As the stock market data is very noisy in nature, it becomes very challenging to glean out a true signal from a mere noise. In this space, there are still a lot of opportunities, as there are myriad of different factors that can be tested. First and foremost inclusion of comprehensive list of technical and fundamental indicators is always solicited.

Further the pre-processing of the features can take many forms and each indiviual form has significant effects on the results. For example; we have just utilized the uniformization technique for this study, however normal standardization procedures or min-max scaling could also be utilized. Similarly, we only predicted the 1-month future returns, a relatively longer term or shorter term predictions may also prove to be more efficient. Thus, we suggest to adopt a more holistic approach for future researchers interested in this field.

Another dimension of the holistic approach may also take into consideration the concept of portfolio building. As already discussed, the application of machine learning methods are closely related to the task of predicting values for financial assets, an important required future contribution should focus on the whole portfolio of stocks. Foe example; the application of machine learning tools to optimize the portfolio building process, such as done in (Markovitz, 1952) provides significant opportunities for future contributions.

Majority of the studies as discussed above are based on technical analysis only which are typically the historical price series. By using only price and volume information as a proxy of a complex market, is a major short-coming of the existing literature. These assumptions are rare to hold in financial markets, as there are multiple factors that may effect the overall movement in stock prices. For example, the stock prices of a company may change its behavior due to changes in economic and company specific information or may be due to changes in the investors psychology or expectations. An interesting future direction will be to incorporate the behavioral factors in the model. There is still a lot of scope to work especially in the context of the stock market of Pakistan. We have included the banking sector with only 40 indicators in this research, however in future it will be interesting to explore more sectors with more indicators. Another very promising direction is the mining of significant feaures on textual information or including the sentiments from social media and news to improve the predictability of stock returns.

Appendices

Table 4: Indicators and their types Pakistan

Indicator	Abbrevaition	Indicator Type
Date	DATE	Date
Stock name	STOCK_ID	Factor
1 months return	Returns	Dependent Variable
Total revenue growth	TRG	Growth indicator
Earnings from continued operations growth	EG	Growth indicator
Net income growth	NIG	Growth indicator
Normalized net income growth	NNIG	Growth indicator
Diluted EPS growth	DEPSG	Growth indicator
Common equity growth	CEG	Growth indicator
Total asset growth	TAG	Growth indicator
Tangible book value growth	TBVG	Growth indicator
Dividend per share growth	DPSG	Growth indicator
Stock price	Stock_price	Momentum and technical analysis
Relative strength index	RSI	Momentum and technical analysis
Log of prices	Log_Price	Momentum and technical analysis
Log returns	Log_returns	Momentum and technical analysis
Relative strength index dividends adjusted	RSI_Div_adjusted	Momentum and technical analysis
Return on assets	ROA	Profitability indicator
Return on equity	ROE	Profitability indicator
Return on common equity	ROCE	Profitability indicator
Selling, general and admin margin	SGAM_	Profitability indicator
Earnings from continued operations margin	EM_	Profitability indicator
Net income margin	NIM	Profitability indicator
Price to diluted EPS ratio	P_Diluted_EPS	Valuation indicator
Price to normalized EPS ratio	P_Normalized_EPS	Valuation indicator
Price to sales ratio	P_Sales	Valuation indicator
Price to book value ratio	P_BV	Valuation indicator
Price to tangible book value ratio	P_TangBV	Valuation indicator
Market cap to total revenues	Mcap_Total_Revenues	Valuation indicator
Log of market cap	Log_Market_Cap	Valuation indicator
Diluted market capital	Dil_Mcap_Outstanding	Valuation indicator
Dividend yield	Dividend_Yield	Valuation indicator
Market capitalization	Market_Cap	Vauation indicator
1 year beta	Beta_Yearly	Volatilities
1 year beta squared	Beta_R_squared	Volatilities
1 year price volatility	Price_volatility_Yearly	Volatilities

Table 1: US Indicators

Column Name	Short Description
stock_id	Stock Identification
Date	Time Period
Advt_12M_Usd	Daily volume aveage USD 12 months
Advt_3M_Usd	Daily volume average USD 3 months
Advt_6M_Usd	Daily volume average USD 6 months
Asset_Turnover	Sales to assets ratio
Bb_Yld	Redemption Yield
Bv	book value of the assets
Capex_Ps_Cf	Capital Expenditure on price to sale cash flow
Capex_Sales	Capital Expenditure margin
Cash_Div_Cf	cash dividends cash flow
Cash_Per_Share	cash per share
Cf_Sales	cash flow per share
Debtequity	debt to equity
Div_Yld	dividend yield
Dps	dividend per share
Ebit_Bv	EBIT on book value
Ebit_Noa	EBIT on non operating asset
Ebit_Oa	EBIT on operating asset
Ebit_Ta	EBIT on total asset
Ebitda_Margin	EBITDA margin
Eps	Earnings per share ratio
Eps_Basic	Basic Earnings per share ratio
Eps_Basic_Gr	Growth in earnings per share
Eps_Contin_Oper	earnings per share due to continued operations
Eps_Dil	diluted earnings per share
Ev	enterprise value
Ev_Ebitda	enterprise value on EBITDA
Fa_Ci	fixed assets on common equity
Fcf	Free Cash Flow
Fcf_Bv	Free cash flow to book value ratio
Fcf_Ce	free cash flow to capital employed ratio
Fcf_Margin	free cash flow to sales ratio
Fcf_Noa	free cash flow to net operating assets coverage
Fcf_Oa	free cash flow to operating assets coverage
Fcf_Ta	free cash flow to total assets coverage
Fcf_Tbv	free cash flow to tangible book value
Fcf_Toa	free cash flow to total operating assets
Fcf_Yld	free cash flow to price
Free_Ps_Cf	free cash flow to price sales ratio
Int_Rev	intangibles to sales ratio
Interest_Expense	interest expense coverage
Mkt_Cap_12M_Usd	avg market cap 12 months USD
Mkt_Cap_3M_Usd	avg market cap 3 months USD
Mkt_Cap_6M_Usd	avg market cap 6 months USD

Table 2: US Indicators

Mom_11M_Usd	momentum 1year minus 1month USD
Mom_5M_Usd	momentum 6months minus 1 months USD
Mom_Sharp_11M_Usd	momentum 1year minus 1month USD by volatility
Mom_Sharp_5M_Usd	momentum 6month - 1month USD by volatility
Nd_Ebitda	debt on Earning before interest tax and depreciation
Net_Debt	net Loan
Net_Debt_Cf	net loan to cash flow
Net_Margin	margin
Netdebtyield	debt yield
Ni	net income
Ni_Avail_Margin	net income margin
Ni_Oa	net income to operating asset ratio
Ni_Toa	net income to total operating asset
Noa	net operating Asset
Oa	operating Asset
Ocf	operating Cash Flow
Ocf_Bv	operating cash flow to bv ratio
Ocf_Ce	operating cash flow to capital employed ratio
Ocf_Margin	Operating Cash Flow margin
Ocf_Noa	operating cash flow to net operating asset ratio
Ocf_Oa	operating cash flow to operating asset ratio
Ocf_Ta	operating cash flow on total assets
Ocf_Tbv	operating Cash flow on TBV
Ocf_Toa	operating Cash Flow on total operating assets
Op_Margin	operating margin
Op_Prt_Margin	net margin growth
Oper_Ps_Net_Cf	cash flow from operations per share
Pb	price to book ratio
Pe	price earnings ratio
Ptx_Mgn	pretax to sales ratio
Recurring_Earning_Total_Assets	Reccuring Earnings on total assets
Return_On_Capital	Return on capital employed
Rev	revenue
Roa	return on assets
Roc	return on capital
Roce	return on capital employed
Roe	return on equity
Sales_Ps	price to sales
Share_Turn_12M	average share turnover 12 months
Share_Turn_3M	average share turnover 3 months
Share_Turn_6M	average share turnover 6 months
Ta	total assets

Table 3: US Indicators

Tev_Less_Mktcap	enterprise value minus market cap
Tot_Debt_Rev	total debt on revenue
Total_Capital	total capital
Total_Debt	total debt
Total_Debt_Capital	debt to capital ratio
Total_Liabilities_Total_Assets	liabilities to asset ratio
Vol1Y_Usd	Returns volatility 1 year
Vol3Y_Usd	Returns volatility 3 years
R1M_Usd	return 1 month
R3M_Usd	return 3 months
R6M_Usd	return 6 months
R12M_Usd	return 12 months

References

Aghabozorgi, S. & Teh, Y. W. (2014). Stock market co-movement assessment using a three-phase clustering method. *Expert Systems with Applications*, *41*(4), 1301–1314.

Aldridge, I. & Avellaneda, M. (2019). Neural networks in finance: Design and performance. *The Journal of Financial Data Science*, *1*(4), 39–62.

Ammann, M., Coqueret, G. & Schade, J.-P. (2016). Characteristics-based portfolio choice with leverage constraints. *Journal of Banking & Finance*, *70*, 23–37.

Ang, A., Hodrick, R. J., Xing, Y. & Zhang, X. (2005). The cross-section of volatility and expected returns. *Journal of Finance, forthcoming*.

Asness, C., Frazzini, A., Gormsen, N. J. & Pedersen, L. H. (2020). Betting against correlation: Testing theories of the low-risk effect. *Journal of Financial Economics*, *135*(3), 629–652.

Asness, C., Frazzini, A., Israel, R., Moskowitz, T. J. & Pedersen, L. H. (2018). Size matters, if you control your junk. *Journal of Financial Economics*, *129*(3), 479–509.

Asness, C. S., Moskowitz, T. J. & Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance*, *68*(3), 929–985.

Astakhov, A., Havranek, T. & Novak, J. (2019). Firm size and stock returns: A quantitative survey. *Journal of Economic Surveys*, *33*(5), 1463–1492.

Atsalakis, G. S. & Valavanis, K. P. (2009). Surveying stock market forecasting techniques–part ii: Soft computing methods. *Expert Systems with applications*, *36*(3), 5932–5941.

Bachelier, L. (1900). Théorie de la spéculation. In *Annales scientifiques de l'école normale supérieure* (Vol. 17, pp. 21–86).

Baker, M., Bradley, B. & Wurgler, J. (2011). Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. *Financial Analysts Journal*, *67*(1), 40–54.

Baker, M., Hoeyer, M. F. & Wurgler, J. (2019). Leverage and the beta anomaly. *Journal of Financial and Quantitative Analysis*, 1–24.

Ballings, M., Van den Poel, D., Hespeels, N. & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, *42*(20), 7046–7056.

Bansal, R. & Viswanathan, S. (1993). No arbitrage and arbitrage pricing: A new approach. *The Journal of Finance*, *48*(4), 1231–1262.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics*, *9*(1), 3–18.

Boloorforoosh, A., Christoffersen, P., Fournier, M. & Gouriéroux, C. (2020). Beta risk in the cross-section of equities. *The Review of Financial Studies*, *33*(9), 4318–4366.

Bouchaud, J.-P., Krueger, P., Landier, A. & Thesmar, D. (2019). Sticky expectations and the profitability anomaly. *The Journal of Finance*, *74*(2), 639–674.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Breiman, L., Friedman, J., Stone, C. J. & Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Bryzgalova, S., Pelger, M. & Zhu, J. (2019). Forest through the trees: Building cross-sections of stock returns. *Available at SSRN 3493458*.

Burrell, P. R. & Folarin, B. O. (1997). The impact of neural networks in finance. *Neural Computing & Applications*, *6*(4), 193–200.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, *52*(1), 57–82.

Cavalcante, R. C., Brasileiro, R. C., Souza, V. L. F., Nóbrega, J. P. & Oliveira, A. (2016). Computational intelligence and financial markets: A survey and future directions. *Expert Syst. Appl.*, *55*, 194-211.

Chen, L., Pelger, M. & Zhu, J. (2019). Deep learning in asset pricing. *Available at SSRN 3350138*.

Chinco, A., Clark-Joseph, A. D. & Ye, M. (2019). Sparse signals in the cross-section of returns. *The Journal of Finance*, *74*(1), 449–492.

Chordia, T., Goyal, A. & Shanken, J. A. (2017). Cross-sectional asset pricing with individual stocks: betas versus characteristics. *Available at SSRN 2549578*.

Cochrane, J. H. (2009). *Asset pricing: Revised edition*. Princeton university press.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of finance*, *66*(4), 1047–1108.

Coqueret, G. & Guida, T. (2020a). *Machine learning for factor investing: R version*. CRC Press.

Coqueret, G. & Guida, T. (2020b). Training trees on tails with applications to portfolio choice. *Annals of Operations Research*, 1–41.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, *2*(4), 303–314.

Daniel, K. & Titman, S. (1997). Evidence on the characteristics of cross sectional variation in stock returns. *the Journal of Finance*, *52*(1), 1–33.

Daniel, K., Titman, S. et al. (2012). Testing factor-model explanations of market anomalies. *Critical Finance Review*, *1*(1), 103–139.

Daniel, K., Titman, S. & Wei, K. J. (2001). Explaining the cross-section of stock returns in japan: factors or characteristics? *The Journal of Finance*, *56*(2), 743–766.

DeMiguel, V., Martin-Utrera, A., Nogales, F. J. & Uppal, R. (2020). A transaction-cost perspective on the multitude of firm characteristics. *The Review of Financial Studies*, *33*(5), 2180–2222.

Du, K.-L. & Swamy, M. N. (2013). *Neural networks and statistical learning*. Springer Science & Business Media.

Dunis, C. L., Likothanassis, S. D., Karathanasopoulos, A. S., Sermpinis, G. S. & Theofilatos, K. A. (2013). A hybrid genetic algorithm–support vector machine approach in the task of forecasting and trading. *Journal of Asset Management*, *14*(1), 52–71.

Eakins, S. G., Stansell, S. R. & Buck, J. F. (1998). Analyzing the nature of institutional demand for common stocks. *Quarterly Journal of Business and Economics*, 33–48.

Esfahanipour, A. & Mousavi, S. (2011). A genetic programming model to generate risk-adjusted technical trading rules in stock markets. *Expert Systems with Applications*, *38*(7), 8438–8445.

Fama, E. F. (1965). The behavior of stock-market prices. *The journal of Business*, *38*(1), 34–105.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of financial economics*, *49*(3), 283–306.

Fama, E. F. & French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, *47*(2), 427–465.

Fama, E. F. & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of*.

Fama, E. F. & French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, *116*(1), 1–22.

Fama, E. F. & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, *81*(3), 607–636.

Feng, G., Giglio, S. & Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance*, *75*(3), 1327–1370.

Feng, G., Polson, N. & Xu, J. (2019). Deep learning in characteristics-sorted factor models. *Available at SSRN 3243683*.

Fischer, T. & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654–669.

Francois, C. (2017). *Deep learning with python.* Manning Publications Company.

Frazzini, A. & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, *111*(1), 1–25.

Freyberger, J., Neuhierl, A. & Weber, M. (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies*, *33*(5), 2326–2377.

Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep learning*. MIT press.

Graham, B. & Dodd, D. L. (1934). *Security analysis. new york: Whittlesey house.* McGraw-Hill Book Company.

Green, J., Hand, J. R. & Zhang, X. F. (2013). The supraview of return predictive signals. *Review of Accounting Studies*, *18*(3), 692–730.

Gu, S., Kelly, B. & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Gu, S., Kelly, B. & Xiu, D. (2021). Autoencoder asset pricing models. *Journal of Econometrics*, *222*(1), 429–450.

Guida, T. & Coqueret, G. (2018). Ensemble learning applied to quant equity: gradient boosting in a multifactor framework. *Big data and machine learning in quantitative investment*, 129–148.

Guliyev, N. J. & Ismailov, V. E. (2018). On the approximation by single hidden layer feedforward neural networks with fixed weights. *Neural Networks*, *98*, 296–304.

Han, Y., He, A., Rapach, D. & Zhou, G. (2019). Firm characteristics and expected stock returns. *Available at SSRN 3185335*.

Harvey, C. R., Liu, Y. & Zhu, H. (2016). … and the cross-section of expected returns. *The Review of Financial Studies*, *29*(1), 5–68.

Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Hjalmarsson, E. & Manchev, P. (2012). Characteristic-based mean-variance portfolio choice. *Journal of Banking & Finance*, *36*(5), 1392–1401.

Hou, K., Xue, C. & Zhang, L. (2020). Replicating anomalies. *The Review of Financial Studies*, *33*(5), 2019–2133.

Hou, K., Xue, C., Zhang, L., Payzan-LeNestour, E., Bossaerts, P., Norli, Ø., … others (n.d.). Digesting anomalies: An investment approach. *Review of Financial Studies*.

Huang, W., Nakamori, Y. & Wang, S.-Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & operations research*, *32*(10), 2513–2522.

Huck, N. (2019). Large data sets and machine learning: Applications to statistical arbitrage. *European Journal of Operational Research*, *278*(1), 330–342.

Iqbal, J. & Brooks, R. (2008, 06). A test of capm on the karachi stock exchange.

Jan, M. N. (2019). *Asset pricing and artificial neural networks: A case of pakistan's equity market* (Unpublished doctoral dissertation). CAPITAL UNIVERSITY.

Jegadeesh, N. & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, *48*(1), 65–91.

Jensen, M. C., Black, F. & Scholes, M. S. (1972). The capital asset pricing model: Some empirical tests.

Jiang, W. (2020). Applications of deep learning in stock market prediction: recent progress. *arXiv preprint arXiv:2003.01859*.

Kelly, B. T., Pruitt, S. & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, *134*(3), 501–524.

Kim. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, *55*(1-2), 307–319.

Kim, S., Korajczyk, R. A. & Neuhierl, A. (2019). Arbitrage portfolios. *Georgia Tech Scheller*

College of Business Research Paper(18-43).

Koijen, R. S., Richmond, R. J. & Yogo, M. (2019). Which investors matter for global equity valuations and expected returns? *Available at SSRN*.

Koijen, R. S. & Yogo, M. (2019). A demand system approach to asset pricing. *Journal of Political Economy*, *127*(4), 1475–1515.

Kozak, S., Nagel, S. & Santosh, S. (2018). Interpreting factor models. *The Journal of Finance*, *73*(3), 1183–1223.

Krauss, C., Do, X. A. & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the s&p 500. *European Journal of Operational Research*, *259*(2), 689–702.

Kremer, P. J., Lee, S., Bogdan, M. & Paterlini, S. (2020). Sparse portfolio selection via the sorted l1-norm. *Journal of Banking & Finance*, *110*, 105687.

Kumar, D. A. & Murugan, S. (2013). Performance analysis of indian stock market index using neural network time series model. In *2013 international conference on pattern recognition, informatics and mobile engineering* (pp. 72–78).

Lee, M.-C. (2009). Using support vector machine with a hybrid feature selection method to the stock trend prediction. *Expert Systems with Applications*, *36*(8), 10896–10904.

Lim, B. & Zohren, S. (2020). Time series forecasting with deep learning: A survey. *arXiv preprint arXiv:2004.13408*.

Luo, J., Subrahmanyam, A. & Titman, S. (2019). Momentum and reversals when overconfident investors underestimate their competition. *The Review of Financial Studies*.

Majumder, D. (2013). Towards an efficient stock market: Empirical evidence from the indian market. *Journal of Policy Modeling*, *35*(4), 572–587.

Malkiel, B. & Fama, E. (1970). Efficient capital markets. *A review of theory and empirical work Journal of Finance 25, 383*, *417*.

Markovitz, H. (1952). Portfolio selection./h' markovitz. *J. of Finance*, *7*(1), 77–91.

Markowitz, H. (1952). Selection, portfolio and diversification, e. *The Journal of Finance*, *7*(1), 77–91.

Martin, I. & Nagel, S. (2019). *Market efficiency in the age of big data* (Tech. Rep.). National Bureau of Economic Research.

Matias, J. M. & Reboredo, J. C. (2012). Forecasting performance of nonlinear models for intraday stock returns. *Journal of Forecasting*, *31*(2), 172–188.

McGee, R. & Olmo, J. (2020). Optimal characteristic portfolios. *Available at SSRN*.

McLean, R. D. & Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, *71*(1), 5–32.

Moritz, B. & Zimmermann, T. (2016). Tree-based conditional portfolio sorts: The relation between past and future stock returns. *Available at SSRN 2740751*.

Murphy, J. J. (1999). *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. Penguin.

Patel, J., Shah, S., Thakkar, P. & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert systems with applications*, *42*(1), 259–268.

PSX. (2021). *Companies Information fianancial reports.* Retrieved 2021-01-30, from `https://dps.psx.com.pk/downloads`

Rapach, D. E., Strauss, J. K., Tu, J. & Zhou, G. (2019). Industry return predictability: A machine learning approach. *The Journal of Financial Data Science*, *1*(3), 9–28.

Rapach, D. E., Strauss, J. K. & Zhou, G. (2013). International stock return predictability: what is the role of the united states? *The Journal of Finance*, *68*(4), 1633–1662.

Rapach, D. E. & Zhou, G. (2020). Time-series and cross-sectional stock return forecasting: New machine learning methods. *Machine Learning for Asset Management: New Developments and Financial Applications*, 1–33.

Rasheed, S., Saood, U. & Alam, W. (2019, 06). Momentum effect in stock market: Empirical evidence from pakistan stock exchange. *Abasyn Journal of Social Sciences*, *12*. doi: 10.34091/AJSS.12.1.02

Recorder, B. (2021). *Companies Information price history.* Retrieved 2021-01-30, from `https://markets.brecorder.com/market-data/company-archives.html`

Richardson, S., Tuna, I. & Wysocki, P. (2010). Accounting anomalies and fundamental analysis: A review of recent research advances. *Journal of Accounting and Economics*, *50*(2-3), 410–454.

Rodríguez-González, A., García-Crespo, Á., Colomo-Palacios, R., Iglesias, F. G. & Gómez-Berbís, J. M. (2011). Cast: Using neural networks to improve trading systems based on technical analysis by means of the rsi financial indicator. *Expert systems with Applications*, *38*(9), 11489–11500.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, *65*(6), 386.

Rowshandel, S. R., Anvary Rostamy, A. A., Noravesh, I. & Darabi, R. (2017). Examination of the predictive power of fama-french five-factor model by the inclusion of skewness coefficient: Evidence of iranian stock market. *International Journal of Finance & Managerial Accounting*, *2*(6), 71–78.

Samuelson, P. A. (2016). Proof that properly anticipated prices fluctuate randomly. In *The world scientific handbook of futures markets* (pp. 25–38). World Scientific.

Schwert, G. W. (2003). Anomalies and market efficiency. *Handbook of the Economics of Finance*, *1*, 939–974.

Sezer, O. B., Gudelek, M. U. & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, *90*, 106181.

Shanaev, S. & Ghimire, B. (2020). Efficient scholars: academic attention and the disappearance of anomalies. *The European Journal of Finance*, 1–27.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, *19*(3), 425–442.

Si, Y.-W. & Yin, J. (2013). Obst-based segmentation approach to financial time series. *Engineering Applications of Artificial Intelligence*, *26*(10), 2581–2596.

Simonian, J., Wu, C., Itano, D. & Narayanam, V. (2019). A machine learning approach to risk factors: A case study using the fama–french–carhart model. *The Journal of Financial Data Science*, *1*(1), 32–44.

Soleymani, F. & Paquet, E. (2020). Financial portfolio optimization with online deep reinforcement learning and restricted stacked autoencoder-deepbreath. *Expert Systems with Applications*, 113456.

Stevens, G. V. (1998). On the inverse of the covariance matrix in portfolio analysis. *The Journal* *of Finance*, *53*(5), 1821–1827.

Tauseef, S. & Nishat, M. (2016). Can investors benefit from momentum trading? evidence from an emerging market.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Tobek, O. & Hronec, M. (2020). Does it pay to follow anomalies research? machine learning approach with international evidence. *Journal of Financial Markets*, 100588.

Uematsu, Y. & Tanaka, S. (2019). High-dimensional macroeconomic forecasting and variable selection via penalized regression. *The Econometrics Journal*, *22*(1), 34–56.

Van Dijk, M. A. (2011). Is size dead? a review of the size effect in equity returns. *Journal of Banking & Finance*, *35*(12), 3263–3274.

Wilder, J. W. (1978). *New concepts in technical trading systems*. Trend Research.

Yoo, P. D., Kim, M. H. & Jan, T. (2005). Machine learning techniques and use of event information for stock market prediction: A survey and evaluation. In *International conference on computational intelligence for modelling, control and automation and international conference on intelligent agents, web technologies and internet commerce (cimca-iawtic'06)* (Vol. 2, pp. 835–841).